

Concentration and Moment Inequalities for Polynomials of Independent Random Variables*

Warren Schudy[†]

Maxim Sviridenko[‡]

Abstract

In this work we design a general method for proving moment inequalities for polynomials of independent random variables. Our method works for a wide range of random variables including Gaussian, Boolean, exponential, Poisson and many others. We apply our method to derive general concentration inequalities for polynomials of independent random variables. We show that our method implies concentration inequalities for some previously open problems, e.g. permanent of random symmetric matrices. We show that our concentration inequality is stronger than the well-known concentration inequality due to Kim and Vu [31]. The main advantage of our method in comparison with the existing ones is a wide range of random variables we can handle and bounds for previously intractable regimes of high degree polynomials and small expectations. On the negative side we show that even for boolean random variables each term in our concentration inequality is tight.

1 Introduction

Concentration and moment inequalities are vital for many applications in Discrete Mathematics, Theoretical Computer Science, Operations Research, Machine Learning and other fields. In the classical setting we have n independent random variables X_1, \dots, X_n and we are interested in a behavior of a function $f(X_1, \dots, X_n)$ of these random variables. Probably, the first concentration inequality with exponential bounds for tails was proven by S. Bernstein [11] who showed that if X_i are random variables that take values $+1$ or -1 with probability $1/2$ (i.e. Rademacher random variables) then

$$\Pr \left[\left| \sum_{i=1}^n X_i \right| \geq \varepsilon n \right] \leq 2e^{-\frac{\varepsilon^2 n}{2(1+\varepsilon/3)}}.$$

More general inequalities known as Chernoff Bounds became part of the mathematical jargon to the extent that many papers in Theoretical Computer Science use them without stating the inequalities. In the last 20 years this area of Probability Theory and related area of mathematics studying the measure concentration has flourished driven by the variety of applications and settings. The surveys and books [22, 17, 13, 39, 34] provide the historical and mathematical background in this area.

The most general and powerful methods known up to date to prove such inequalities is Ledoux's entropy method [34] and the famous Talagrand's isoperimetric inequality [45]. Yet as was noticed by Vu [49] these methods and corresponding inequalities work well only when the Lipschitz coefficients of the function $f(X_1, \dots, X_n)$ are relatively small. The standard example showing the weakness

*An extended 10-page abstract of this paper appeared in Proceedings of the Symposium on Discrete Algorithms (SODA2012)

[†]IBM T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, NY 10598. wjschudy@us.ibm.com

[‡]University of Warwick, sviri@dcs.warwick.ac.uk

of such methods is the number of triangles in random graphs $G(n, p)$. Until the concentration inequality due to Kim and Vu [31] no non-trivial concentration of this function about its mean was known.

Kim and Vu [31] introduced the notion of average Lipschitz coefficients based on the partial derivatives of a polynomial evaluated at the point $(\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$ (in the multilinear case). These new parameters enabled them to prove a concentration inequality for polynomials of boolean random variables. This inequality has been applied to the problem of approximately counting triangles in (e.g.) a social network by sampling the edges [47, 48], to average-case correlation clustering [37], and to a variety of other applications [49]. The original inequality from [31] was tightened and generalized in [49] to handle arbitrary random variables in the interval $[0, 1]$. Yet the inequality from [49] did not work well for high degree polynomials and for random variables $f(X_1, \dots, X_n)$ with small expectation. The follow up work by Vu [50] handles the case of polynomials with small expectation and extremely small smoothness parameters.

On the other side the concentration of polynomials of Gaussian and Rademacher random variables has long been a subject of interest in Probability Theory. The moment and concentration inequalities for polynomials of centered Gaussians are known as Hypercontractivity Inequalities [42, 24]. We discuss various inequalities known in this setting and their connection to our results in Section 1.5. Recently, the Hypercontractivity Inequalities and their “anti-concentration” counterparts found many applications in Theoretical Computer Science and Machine Learning [10, 18, 19, 20, 21, 29].

The above motivated us to study the moment and concentration inequalities for polynomials of independent random variables. We design a general method that works for a wide range of random variables including Gaussian, Boolean, exponential, Poisson and many others (see Section 7 for more examples). We show that our method implies concentration inequalities for some previously open problems, e.g. permanent of random symmetric matrices. We also show that our main concentration inequality is stronger than the well-known concentration inequality due to Kim and Vu [31]. On the negative side we show that even for boolean random variables each term in our concentration inequality is tight.

1.1 Our Results

For a cleaner exposition we first describe our results in the restricted setting of multilinear polynomials with non-negative coefficients. We are given a hypergraph $H = (\mathcal{V}(H), \mathcal{H}(H))$ consisting of a set $\mathcal{V}(H) = \{1, 2, \dots, n\} = [n]$ of vertices and a set $\mathcal{H}(H)$ of hyperedges. A hyperedge h is a set $h \subseteq \mathcal{V}(H)$ of $|h| \leq q$ vertices. We are also given a non-negative weight w_h for each $h \in \mathcal{H}(H)$. For each such weighted hypergraph and real-valued weight w_h for its hyperedges, we define a polynomial

$$f(x) = \sum_{h \in \mathcal{H}(H)} w_h \prod_{v \in h} x_v. \quad (1.1)$$

Our smoothness parameters were strongly motivated by the average partial derivatives introduced by Kim and Vu [31, 49]. For any $y \in \mathbb{R}^n$, hypergraph H , nonnegative weights w , and $h_0 \subseteq \mathcal{V}(H)$ let

$$\mu(y, H, w, h_0) = \sum_{h \in \mathcal{H}(H) \mid h \supseteq h_0} w_h \prod_{v \in h \setminus h_0} |y_v|.$$

Note that h_0 need not be a hyperedge of H and may even be the empty set. Also note that $\mu(y, H, w, h_0)$ is equal to the $|h_0|$ -th partial derivative of polynomial $f(x)$ with respect to each variable x_v for $v \in h_0$, evaluated at the point $x = y$ if $y \in \mathbb{R}_+^n$. For a given collection of independent

random variables $Y = (Y_1, \dots, Y_n)$, hypergraph H , integer $r \geq 0$ and nonnegative weights w , we define

$$\mu_r = \mu_r(H, w) = \max_{h_0 \subseteq [n]: |h_0|=r} \mathbb{E} [\mu(Y, H, w, h_0)] = \max_{h_0 \subseteq [n]: |h_0|=r} \left(\sum_{h \in \mathcal{H}(H) | h \supseteq h_0} w_h \prod_{v \in h \setminus h_0} \mathbb{E} [|Y_v|] \right),$$

where we used the independence of random variables Y_v in the last equality. Sometimes we will also use the notation $\mu_r(f) = \mu_r(H, w)$. Note that when the Y_v are non-negative μ_r is equal to the maximal expected partial derivative of order r of the polynomial $f(x)$, which was the parameter used in the Kim-Vu concentration inequalities [31, 49].

Our concentration inequalities will hold for a general class of independent random variables including most classical ones.

Definition 1.1 *A random variable Z is called moment bounded with parameter $L > 0$ if for any integer $i \geq 1$,*

$$\mathbb{E} [|Z|^i] \leq i \cdot L \cdot \mathbb{E} [|Z|^{i-1}].$$

Roughly speaking a random variable Z is moment bounded with parameter L if $\mathbb{E} [|Z|] \leq L$ and the tails of its distribution decay no slower than an exponentially distributed random variable's tails do. Indeed note that Definition 1.1 implies that any moment bounded random variable Z satisfies $\mathbb{E} [|Z|^i] \leq L^i i!$. In Section 7 we show that three large classes of random variables are moment bounded: bounded, continuous log-concave [8, 5] and discrete log-concave [5]. For example the Poisson, binomial, geometric, normal (i.e. Gaussian), and exponential distributions are all moment bounded.

We prove the following:

Theorem 1.2 *We are given n independent moment bounded random variables $Y = (Y_1, \dots, Y_n)$ with the same parameter L . We are given a multilinear polynomial $f(x)$ with nonnegative coefficients of total power¹ q . Let $f(Y) = f(Y_1, \dots, Y_n)$ then*

$$\Pr [|f(Y) - \mathbb{E}[f(Y)]| \geq \lambda] \leq e^2 \cdot \max \left\{ \max_{r=1, \dots, q} e^{-\frac{\lambda^2}{\mu_0 \mu_r \cdot L^r \cdot R^q}}, \max_{r=1, \dots, q} e^{-\left(\frac{\lambda}{\mu_r \cdot L^r \cdot R^q}\right)^{1/r}} \right\},$$

where $R \geq 1$ is some absolute constant.

We also show that Theorem 1.2 is the best possible bound as a function of these parameters, up to logarithms in the exponent and dependence of the constants on the total power q . This lower bound holds even for the well-studied special case where the random variables take the values 0 and 1 only, which we show in Section 7 to be moment bounded with parameter 1.

Theorem 1.3 *For any $q \in \mathbb{N}$, real numbers $\mu_0^*, \mu_1^*, \dots, \mu_q^* > 0$ and $\lambda > 0$ there exist independent 0/1 random variables $X = X_1, \dots, X_n$ and a polynomial $f(x)$ of power q such that $\mu_i(f) \leq \mu_i^*$ for all $0 \leq i \leq q$ and*

$$\Pr [f(X) \geq \mathbb{E} [f(X)] + \lambda] \geq \max_{r=1, \dots, q} \max \left\{ e^{-\left(\frac{\lambda^2}{\mu_0^* \mu_r^*} + 1\right) \log C}, e^{-\left(\left(\frac{\lambda}{\mu_r^*}\right)^{1/r} + 1\right) \log C} \right\} \quad (1.2)$$

where $C = c_0 \Lambda_1^{c_1} \Lambda_2^{c_2} \Lambda_3^{c_3}$, c_0, c_1, c_2 and c_3 are absolute constants, $\Lambda_1 = \max_{0 \leq i, j \leq q} (\mu_i^* / \mu_j^*)^q$, $\Lambda_2 = \max_{1 \leq i \leq q} \lambda / \mu_i^*$, and $\Lambda_3 = q^q$.

¹We reserve the more traditional terminology of “degree” for the number of neighbors of a vertex in a hypergraph.

We generalize Theorem 1.2 in two ways. Firstly, we allow negative coefficients. Secondly, we remove the restriction for a polynomial to be multilinear, instead allowing each monomial to have total power at most q and maximal power of each variable at most Γ . For example $X_1^2 X_2^4 X_3^1$ has total power $q = 7$ and maximal variable power $\Gamma = 4$ and the multilinear case is when maximal power is $\Gamma = 1$. We defer the formal definition of general polynomials and the appropriate generalization of μ_r to Section 1.4.

Our main result in this paper is the following:

Theorem 1.4 *We are given n independent moment bounded random variables $Y = (Y_1, \dots, Y_n)$ with the same parameter L . We are given a general polynomial $f(x)$ of total power q and maximal variable power Γ . Let $f(Y) = f(Y_1, \dots, Y_n)$ then*

$$Pr[|f(Y) - \mathbb{E}[f(Y)]| \geq \lambda] \leq e^2 \cdot \max \left\{ \max_{r=1, \dots, q} e^{-\frac{\lambda^2}{\mu_0 \mu_r \cdot L^r \cdot \Gamma^r \cdot R^q}}, \max_{r=1, \dots, q} e^{-\left(\frac{\lambda}{\mu_r \cdot L^r \cdot \Gamma^r \cdot R^q}\right)^{1/r}} \right\},$$

where $R \geq 1$ is some absolute constant.

For large power polynomials the concentration bounds in the Theorem 1.4 may not provide interesting concentration bounds due to the term R^q in the exponent, yet we believe that the moment computation method developed in this paper is useful even in this setting. We show two specific examples when our method works. Our first example is a concentration inequality for permanents of random matrices. The anti-concentration counterpart was recently studied by Aaronson and Arkhipov [1] in the Gaussian setting and by Tao and Vu [46] in the setting with Rademacher random variables.

Theorem 1.5 *We are given $n \times n$ matrix A with random entries Y_{ij} which are independent moment bounded random variables with parameter $L = 1$ and $\mathbb{E}[Y_{ij}] = 0$. Let $P(A)$ be the permanent of the matrix A then*

$$Pr[|P(A)| \geq t\sqrt{n!}] \leq \max \left\{ e^{-n}, e^2 \cdot e^{-c \cdot t^2/n} \right\}$$

for some absolute constant $c > 0$ and parameter $t > 0$.

Our next example is an analogous Theorem for the permanent of a random symmetric matrix.

Theorem 1.6 *We are given $n \times n$ symmetric matrix A with random entries Y_{ij} which are independent moment bounded random variables for all pairs (i, j) with $i \leq j$ with parameter $L = 1$ and $\mathbb{E}[Y_{ij}] = 0$. Let $P(A)$ be the permanent of the matrix A then*

$$Pr[|P(A)| \geq t\sqrt{n!}] \leq \max \left\{ e^{-n}, e^2 \cdot e^{-c \cdot t^2/n} \right\}$$

for some absolute constant $c > 0$ and parameter $t > 0$.

Note that the above concentration inequalities can be easily derived from the Hypercontractivity Inequality in the special case of Gaussian and Rademacher random variables (Theorem 1.9).

1.2 Applications in Randomized Rounding for Mathematical Programming Problems

As we noted all current methods to prove concentration bounds for polynomials do not work well for high power polynomials. Another feature that makes current concentration methods fail is low

expectation. One application where such concentration bounds could be applied is in design and analysis of randomized rounding algorithms for non-linear mathematical programming problems.

Many real-life optimization problems can be formulated using integer programming which is well-known to be computationally intractable (NP-hard). One way to solve such a problem both in theory and practice is to consider a linear programming relaxation, solve it using one of the standard methods and use the fractional optimal solution as a guidance in finding an integral solution of good quality. The seminal paper of Raghavan and Thompson [43] suggested to round each boolean variable to one with probability x_i^* and to zero with probability $1 - x_i^*$ independently at random where x^* is the optimal fractional solution. The analysis of such algorithms is based on applying Chernoff Bounds to each constraint of the integer program separately and then applying a union bound over all the constraints. Such a method proved to be useful for a wide range of models and led to approximation algorithms that still have best known performance guarantees today.

A natural generalization of this framework is to apply it to non-linear optimization models. Many such problems are still computationally tractable if we replace the constraint that variables must be boolean $x_i \in \{0, 1\}$ with continuous constraints $0 \leq x_i^* \leq 1$, e.g. quadratic convex constraints. There are many real-life optimization problems with constraints and objective functions modeled in such a way, e.g. we would like to optimize a congestion for a group of edges in a multi-commodity flow problem in a "fair" way, i.e. we don't want to have one edge to get significantly higher congestion than the other. The standard way to ensure that in practice is to optimize (or constrain) sum of squared congestions over the edges in a that group. The constraints generated this way are convex quadratic constraints and continuous optimization problems with such constraints are polynomially solvable.

To analyze the randomized rounding framework for such mathematical programming models one needs to apply concentration inequality to each non-linear constraint. If the size of the group of edges for which we are trying to optimize the total congestion in a fair way is sub-logarithmic and each edge in the fractional solution has a constant congestion (a situation quite natural from application viewpoint) then our concentration inequalities would be the only available tool to analyze such an algorithm.

1.3 Sketch of Our Methods

Most concentration results for non-negative random variables are proven using Markov's inequality as follows:

$$\Pr[Z \geq \lambda] = \Pr[g(Z) \geq g(\lambda)] \leq \frac{\mathbb{E}[g(Z)]}{g(\lambda)} \quad (1.3)$$

where Z is the random variable that we are trying to show concentration of and g is either $g(z) = z^k$ for some positive even integer k , $g(z) = e^{tz}$ for some real $t > 0$, or some other non-negative increasing function g . One then computes an upper bound on either the k th moment $\mathbb{E}[g(Z)] = \mathbb{E}[Z^k]$ or the moment generating function $\mathbb{E}[g(Z)] = \mathbb{E}[e^{tZ}]$. Chernoff bounds are proven using moment generating functions, so it would be most natural to use moment generating functions to prove our bounds as well. Unfortunately the tails of the distribution of polynomials can be sufficiently large to make the moment generating function $\mathbb{E}[e^{tZ}]$ infinite for all $t > 0$. Kim and Vu worked around this issue by applying (1.3) not to the polynomial itself but to various auxiliary random variables with better behaved tails. Unfortunately a union bound over these auxiliary variables introduced an extraneous factor logarithmic in the number of variables into their bounds (see Section 1.5 for a comparison of our results to theirs). We avoid this issue by computing moments instead of the

moment generating function.

We now give an instructive half-page bound on the second moment of a multilinear polynomial $f(X) = \sum_{h \in \mathcal{H}} w_h \prod_{v \in h} X_v$ where all $\mathbb{E}[X_v] = 0$, X_v are moment bounded with parameter L , and all $h \in \mathcal{H}$ have $|h| = q$ and $w_h \geq 0$. Using definitions, linearity of expectation, and independence we get

$$\begin{aligned} \mathbb{E}[f(X)^2] &= \mathbb{E}\left[\sum_{h_1 \in \mathcal{H}} \sum_{h_2 \in \mathcal{H}} w_{h_1} w_{h_2} \left(\prod_{v \in h_1} Y_v\right) \left(\prod_{v \in h_2} Y_v\right)\right] \\ &= \sum_{h_1 \in \mathcal{H}} \sum_{h_2 \in \mathcal{H}} w_{h_1} w_{h_2} \prod_{v \in (h_1 \cup h_2)} \mathbb{E}[Y_v^{d_v}] \end{aligned} \quad (1.4)$$

where $d_v \in \{1, 2\}$ is the number of $h_i \ni v$. Now if $d_v = 1$ for any v we have $\mathbb{E}[Y_v^{d_v}] = \mathbb{E}[Y_v] = 0$, so the only non-zero terms of the sum (1.4) are when $h_1 = h_2$. We therefore get

$$\begin{aligned} \mathbb{E}[f(X)^2] &= \sum_{h \in \mathcal{H}} w_h w_h \prod_{v \in h} \mathbb{E}[Y_v^2] \\ &\leq \sum_{h \in \mathcal{H}} \mu_q w_h \prod_{v \in h} (2L \mathbb{E}[|Y_v|]) \\ &= (2L)^q \mu_q \sum_{h \in \mathcal{H}} w_h \prod_{v \in h} \mathbb{E}[|Y_v|] \\ &= (2L)^q \mu_q \mu_0. \end{aligned} \quad (1.5)$$

where we used the fact that $\mathbb{E}[Y_v^2] \leq 2L \mathbb{E}[|Y_v|]$ from moment boundedness and $w_h \leq \max_h w_h = \mu_q$ from the definition of μ_q . Combining (1.5) with Markov's inequality (1.3) yields

$$\Pr[|f(X)| \geq \lambda] \leq \left(\frac{\lambda^2}{(2L)^q \mu_q \mu_0}\right)^{-1}.$$

which is comparable to the $e^{-\lambda^2/(\mu_0 \mu_q (RL)^q)}$ term in Theorem 1.2 for small λ . In order to get exponentially better bounds for larger λ we will compute higher moments.

Now we outline what we do differently to handle higher moments and general polynomials.

The first step is to express polynomial f over variables Y_v as a sum of polynomials $g^{(1)}, \dots, g^{(m)}$ over variables $Y_v^\tau - \mathbb{E}[Y_v^\tau]$ for various $1 \leq \tau \leq q$. The main task is bounding the moments of each of these polynomials. We later combine these bounds to get a bound on the moment of f . Each of the centered polynomials has $\mathbb{E}[Y_v^\tau - \mathbb{E}[Y_v^\tau]] = 0$, which takes the place of the $\mathbb{E}[Y_v] = 0$ in the above special case. We also ensure that each $g^{(i)}$ has non-negative weights.

Bounding moments of some g_i begins by expanding $\mathbb{E}[g_i^k]$ similar to (1.4) with a sum over h_1, \dots, h_k . As before only terms of the sum where every vertex v occurs in $d_v \geq 2$ different hyperedges are non-zero, but this is no longer equivalent to the simple condition $h_1 = h_2$.

We find it helpful to separate the structure of the hyperedges h_1, \dots, h_k from the identity of the variables involved. We therefore generate h_1, \dots, h_k by composing two processes: first generate h_1, \dots, h_k over vertex set $[\ell]$ for every $\ell \geq 1$ and then consider every possible embedding of those artificial vertices into the vertex set $[n]$. For a fixed sequence of hyperedges over vertex set $[\ell]$ we do arguments analogous to (1.5) to get a product of various μ_i and L . This bound is a function of the number of connected components c in h_1, \dots, h_k . Finally we do some combinatorics to prove

a *counting lemma* on the number of possible h_1, \dots, h_k with vertex set $[\ell]$ with all degrees at least two and c connected components.

One additional complication is that we need to use moment boundedness to bound moments of order much larger than the second moments $\mathbb{E}[Y_v^2] \leq 2L\mathbb{E}[|Y_v|]$ we used in the above special case. If we treated the factor that replaces that 2 as a constant that would make the constant R in our final bounds linear in q instead of an absolute constant. Fortunately these extra factors are small for *most* of the possible h_1, \dots, h_k , which enables our counting lemma to absorb these extra factors.

Our lower bounds are based on lower-bounding the concentration of certain concrete polynomials. It is well known that Chernoff bounds are essentially tight, i.e. a sum of n i.i.d. $0/M$ random variables each with expected value μ/n has probability roughly $e^{-\lambda^2/(2\mu M)}$ of exceeding its mean by $\lambda \leq \mu$. Our lower bound of $e^{-\tilde{O}(\lambda^2/(\mu_0\mu_r))}$ follows from a degree q polynomial that acts like this linear polynomial with $M = \mu_r$ and $\mu = \mu_0$. The idea behind the lower bound corresponding to $e^{-(\lambda/\mu_r)^{1/r}}$ is the fact that $\Pr[(\sum_i X_i)^r \geq \lambda] = \Pr[\sum_i X_i \geq \lambda^{1/r}] = e^{-\tilde{\Theta}(\lambda^{1/r})}$ where $\sum_i X_i$ is binomially distributed with mean 1. Our lower bound does similar arguments with a multilinearized version of $(\sum_i X_i)^r$.

1.4 Definitions

We now state the generalizations of the notations given in the introduction for general polynomials.

A *powered hypergraph* H consists of a set $\mathcal{V}(H)$ of vertices and a set $\mathcal{H}(H)$ of powered hyperedges. A *powered hyperedge* h consists of a set $\mathcal{V}(h) \subseteq \mathcal{V}(H)$ of $|\mathcal{V}(h)| = \eta(h)$ vertices and an $\eta(h)$ -element *power vector* $\tau(h)$ with one strictly positive integer component $\tau(h)_v = \tau_{hv}$ per vertex $v \in \mathcal{V}(h)$. We will hereafter omit the “powered” from “powered hypergraph” and “powered hyperedge” since we have no need to refer to the basic hypergraphs used in the introduction. For any powered hyperedge h we let $q(h) = \sum_{v \in \mathcal{V}(h)} \tau_{hv}$. For each such powered hypergraph H and real-valued weights w_h for its hyperedges, we define a polynomial

$$f(x) = \sum_{h \in \mathcal{H}(H)} w_h \prod_{v \in \mathcal{V}(h)} x_v^{\tau_{hv}}. \quad (1.6)$$

The hyperedge h corresponds to a monomial $\prod_{v \in h} x_v^{\tau_{hv}}$. The parameters $q(h)$ and $\eta(h)$ will be called the *total power* and *cardinality* of the hyperedge h (or monomial corresponding to h). Let $\Gamma = \max_{h \in \mathcal{H}(H), v \in h} \tau_{hv}$ be the maximal power of a variable in polynomial $f(x)$, e.g. $\Gamma = 1$ for multilinear polynomials. We assume, by convention, that $\prod_{i \in \emptyset} x_i = 1$. Since the variables in our polynomials are indexed by vertices in our hypergraphs we use the terms “variable” and “vertex” interchangeably.

For powered hyperedges h_1 and h_2 (not necessarily hyperedges of a hypergraph) we write $h_1 \succeq h_2$ if $\mathcal{V}(h_1) \supseteq \mathcal{V}(h_2)$ and $\tau_{h_1 v} = \tau_{h_2 v}$ for all $v \in \mathcal{V}(h_2)$. In the context of hypergraph H with vertex set $[n]$ clear from context, for a given collection of independent random variables $Y = (Y_1, \dots, Y_n)$, integer $r \geq 0$ and weights w we define

$$\mu_r(w, Y) = \max_{h_0 | \mathcal{V}(h_0) \subseteq \mathcal{V}(H), q(h_0)=r} \left(\sum_{h \in \mathcal{H}(H) | h \succeq h_0} |w_h| \prod_{v \in \mathcal{V}(h) \setminus \mathcal{V}(h_0)} \mathbb{E}[|Y_v^{\tau_{hv}}|] \right) \quad (1.7)$$

where h_0 ranges over all possible powered hyperedges with vertices from $[n]$ with total power $q(h_0) = r$. The cardinality of h_0 is not explicitly restricted but it cannot exceed r since the powers $\tau_{h_0 v}$ are strictly positive integers summing to $q(h_0)$. We will sometimes write $\mu_r(f, Y)$ for

polynomial $f(Y)$ instead of $\mu_r(w, Y)$ to emphasize the dependence on polynomial $f(Y)$. If we write $\mu_r(f)$ for a polynomial f this means $\mu_r(w, Y)$ for the weight function w and random variable vector Y corresponding to f as in (1.6). If the polynomial is clear from context we write simply μ_r . In the special case that all coefficients are non-negative μ_r is upper bounded by the maximal expected partial derivative of order r of the polynomial $f(x)$, and this bound is loose for two reasons. First we do not have multipliers that depend on powers that are present in derivatives. Second we throw away some positive terms that are present in derivatives since we enforce $\tau_{hv} = \tau_{h_0v}$ for all $v \in \mathcal{V}(h_0)$, whereas derivatives would consider all h with $\tau_{hv} \geq \tau_{h_0v}$. For example for the polynomial $Y_0^3 Y_1^3 + Y_1^2$ of non-negative random variables we have $\mu_2 = 1$ while the maximal expected second partial derivative is equal to

$$\max \{2 + 6\mathbb{E}[Y_0^3] \mathbb{E}[Y_1], 6\mathbb{E}[Y_0] \mathbb{E}[Y_1^3], 9\mathbb{E}[Y_0^2] \mathbb{E}[Y_1^2]\}.$$

Overall, our definition of smoothness is a bit tighter (although less natural) than the partial derivatives used in [31], [49] that inspired it. We decided to use it since it naturally arises in our analysis.

1.5 Comparison with Known Concentration Inequalities

There are many concentration inequalities dealing with the case when we are interested in a sum of weakly dependent random variables. The paper [26] provides a good survey and comparison of various inequalities for that setting. Below we will survey only known concentration inequalities for the case of polynomials of independent random variables. The previous works were dealing either with the case of boolean random variables, variables distributed in the interval $[0, 1]$, Gaussian random variables or log-concave random variables.

1.5.1 Comparing with the Kim-Vu inequality

Probably the most famous concentration inequality for polynomials is due to Kim and Vu [31] published in 2000. There are many variants, extensions and equivalent formulations of that inequality. We consider a variant from the survey paper by Vu [49] (Theorem 4.2 in Section 4.2).

Theorem 1.7 (Kim-Vu Concentration Inequality) *Consider a polynomial $f(Y) = f(Y_1, \dots, Y_n)$ with coefficients in the interval $[0, 1]$. We denote $\partial_A f(Y)$ a polynomial obtained from $f(Y)$ by taking partial derivatives with respect to A where A is a multiset of indices probably with repetitions. Let Y_1, \dots, Y_n be independent random variables with arbitrary distributions on the interval $[0, 1]$. Let q be the degree of polynomial $f(Y)$ and $\mathbb{E}_j[f(Y)] = \max_{|A| \geq j} \mathbb{E}[\partial_A f(Y)]$. Assume we are given an integer $q' \leq q$ and a collection of positive numbers $\mathcal{E}_0 \geq \mathcal{E}_1 \geq \dots \geq \mathcal{E}_{q'} = 1$ and λ satisfying*

1. $\mathcal{E}_j \geq \mathbb{E}_j[f(Y)]$ for $j = 0, \dots, q'$;
2. $\mathcal{E}_j / \mathcal{E}_{j+1} \geq \lambda + 4j \log n$ for $j = 0, \dots, q' - 1$;

then the following holds

$$\Pr[|f(Y) - \mathbb{E}[f(Y)]| \geq c_q \sqrt{\lambda \mathcal{E}_0 \mathcal{E}_1}] \leq d_q e^{-\lambda/4}$$

where $c_q \approx q^{q/2}$ and $d_q = 2^{q+1} - 2$ (see precise definitions in [49]).

This stronger version of the original Kim-Vu inequality [31] has dependence on parameter q' which could be helpful for some applications (see discussion in [49]). We compare below our inequality with the inequality in Theorem 1.7 when $q' = q$ which includes the original Kim-Vu inequality [31]

and is the most relevant variant in terms of various applications (our inequality does not seem to be comparable with the general version of the Theorem 1.7).

Re-writing our inequality from Theorem 1.4 in the same form we could derive

$$Pr \left[|f(Y) - \mathbb{E}[f(Y)]| \geq \max_{r=1,\dots,q} \max \left\{ \sqrt{\tau \mu_0 \mu_r \cdot L^r \cdot \Gamma^r \cdot R^q}, \tau^r \mu_r \cdot L^r \cdot \Gamma^r \cdot R^q \right\} \right] \leq e^2 \cdot e^{-\tau}$$

instead of the bound in the Theorem 1.4 for any $\tau > 0$. Using the properties of bounds \mathcal{E}_j we derive $\mathcal{E}_0 \geq \lambda^j \mathbb{E}_j[f(Y)]$ and $\mathcal{E}_1 \geq \lambda^{j-1} \mathbb{E}_j[f(Y)]$. In addition, as we already noticed, our definition of smoothness is tighter than the one based on partial derivatives, i.e. $\mathcal{E}_j \geq \mathbb{E}_j[f(Y)] \geq \mu_j$. Therefore,

$$\sqrt{\lambda \mathcal{E}_0 \mathcal{E}_1} \geq \max_{r=1,\dots,q} \max \left\{ \sqrt{\lambda \mu_0 \mu_r}, \lambda^r \mu_r \right\}.$$

Choosing $\tau = \lambda/4$, we obtain that the concentration inequality of Theorem 1.4 implies the inequality from Theorem 1.7 (we don't explicitly specify the relationship between our absolute constant R and constants used in the definition of c_q and d_q). We list below the various ways our inequality generalizes or tightens the inequality from Theorem 1.7.

1. The bounds in our inequality do not depend on the total number of random variables n while all variants of the Kim-Vu inequality have this dependence due to the usage of the union bound in their proof.
2. Our inequality covers a much wider range of random variables, including most commonly used ones not just the variables distributed in the interval $[0, 1]$.
3. Our definition of smoothness while being related to (and strongly motivated by) the smoothness based on partial derivatives is tighter and for some applications involving polynomials with large Γ will provide a better concentration bound.
4. Our bounds have a better dependence on the degree of the polynomials. We also introduce a parameter Γ that is a maximal power of a variable in a polynomial which leads to substantially tighter bounds for the most important special case of multilinear polynomials.

Another concentration inequality that appeared in the literature is due to Boucheron et al. [14] (Section 10).

Theorem 1.8 *Consider a multilinear degree q polynomial $f(Y) = f(Y_1, \dots, Y_n)$ of the independent boolean random variables Y_1, \dots, Y_n . Then*

$$Pr[f(Y) \geq \mathbb{E}[f(Y)] + \lambda] \leq e^{-\frac{1}{R \cdot q}} \max \left\{ \max_{r=1,\dots,q} e^{-\left(\frac{\lambda^2}{16q^2 \mu_0 \mu_r}\right)^{1/r}}, \max_{r=1,\dots,q} e^{-\left(\frac{\lambda}{4q \mu_r}\right)^{1/r}} \right\}$$

for some absolute constant $R > 0$.

The second term in the maximum looks very similar to ours in the Theorem 1.4 but the first term is substantially higher due to the power $1/r$. Also their inequality does not seem to generalize to general class of random variables considered in Theorem 1.4. Note that the Theorem 1.8 is just a corollary of a moment inequality proved for much more general functions than polynomials.

1.5.2 Gaussian and Rademacher Random Variables

Another class of known concentration inequalities deals with the case when random variables are either centered (or zero mean) Gaussians or variables that have value $+1$ or -1 with probability $1/2$ (such random variables are often called Rademacher random variables). The history of moment and concentration inequalities in this setting is quite rich, we refer the reader to the Lecture 16 in Ryan O'Donnell Lecture Notes on Boolean Analysis [42] or the book by S. Janson [24] (Sections V and VI). We will call the moment and corresponding concentration inequalities the Hypercontractivity Inequalities for the formal proofs see Theorems 6.7 and 6.12 in [24].

Theorem 1.9 (Hypercontractivity Concentration Inequality) *Consider a degree q polynomial $f(Y) = f(Y_1, \dots, Y_n)$ of independent centered Gaussian or Rademacher random variables Y_1, \dots, Y_n . Then*

$$\Pr[|f(Y) - \mathbb{E}[f(Y)]| \geq \lambda] \leq e^2 \cdot e^{-\left(\frac{\lambda^2}{R \cdot \text{Var}[f(Y)]}\right)^{1/q}},$$

where $\text{Var}[f(Y)]$ is the variance of the random variable $f(Y)$ and $R > 0$ is an absolute constant.

It is well-known that functions of Gaussian random variables are better concentrated around their mean than for example functions of Boolean random variables even in such a simple case as a sum of independent random variables. Therefore, in general we cannot expect to match the bound of Theorem 1.9 in the setting of moment bounded random variables. Nevertheless, if $M_2 \approx \max_{r \in [q]} \mu_0 \mu_r$ (e.g. it happens when power $q = O(1)$, the polynomial is multilinear, $\mu_r = O(1)$ for $r \in [q-1]$ and $w_h \in \{0, 1\}$ for all hyperedges h) and $\lambda \leq \mu_0$ then Theorem 1.4 provides a better concentration bound even in this setting.

An interesting concentration inequality for degree q polynomials of centered Gaussian random variables was recently proven by R. Latala [32]. This inequality generalizes the previously known inequalities for the case when $q = 2$ [28]. The papers by Major [36] and Lehec [35] simplify and explain Latala's proof. Latala uses certain smoothness parameters that seem to be natural only in the setting of continuous random variables. We do not see the way to define similar smoothness parameters in the setting of general moment bounded (or even boolean) random variables.

1.5.3 Log-Concave Random Variables

We define log-concave random variables in the Section 7 and give many examples of such variables. Latala and Lochowski [33] consider the setting with non-negative log-concave random variables and multi-linear polynomials. Recently, Adamczak and Latala [3] considered symmetric log-concave random variables and polynomials of degree at most three and symmetric exponential random variables (or variables having Laplace distribution) for polynomials of arbitrary degree. The main drawback of their approach in [33] is that they estimate tails of random variables instead of estimating the deviation from the mean which is required in most applications. We can show that their smoothness parameters can be derived from ours μ_r (and the tail bounds) in the case of exponential random variables or any random variables that are tight for our moment boundness condition, i.e. $\mathbb{E}[|X|^i] \approx iL\mathbb{E}[|X|^{i-1}]$.

1.6 Other Concentration Inequalities for Polynomials

Another line of attack on understanding the concentration of polynomials is to use the structure of polynomials and some smoothness parameters analogous to the partial derivatives or our parameters μ_r . Many of these known inequalities provide tight upper and lower bounds for moments but

involve hard to estimate smoothness parameters. In the case of Gaussian random variables tight concentration bounds for polynomials of independent random variables were obtained by Hanson and Wright [28] for $q = 2$ and Borell [12], Arcones and Gine [7] for $q \geq 3$. In the case of Rademacher random variables analogous results but based on different methods were obtained by Talagrand [44] for $q = 2$ and Boucheron et al. [14] for $q \geq 3$. Adamczak [2] proved a concentration inequality for general functions of a general class of random variables. All these results except [28] and [44] (i.e. the case of quadratic polynomials) use parameters that involve expectations of suprema of certain empirical processes that are in general not easy to estimate which limits applicability of these inequalities.

Another interesting class of inequalities was obtained by using the so-called "needle decomposition method" in the field of Geometric Functional Analysis. It is a rich research area and we refer the reader to the survey paper by Nazarov, Sodin and Volberg [41]. An interesting moment inequality which seem to generalize and tighten many previously known inequalities in this area was shown by Carbery and Wright [16] (see Theorem 7). It implies the following concentration inequality via application of Markov's inequality

Theorem 1.10 *Consider a degree q polynomial $f(Y) = f(Y_1, \dots, Y_n)$. Assume that random variables Y_1, \dots, Y_n are distributed according to some log-concave measure in R^n (i.e. they are not necessarily independent). Then*

$$Pr[|f(Y) - \mathbb{E}[f(Y)]| \geq \lambda] \leq e^2 \cdot e^{-\left(\frac{\lambda}{R\sqrt{\text{Var}[f(Y)]}}\right)^{1/q}}$$

for some absolute constant $R > 0$.

On one side this inequality is extremely general and allows to study such processes as sampling a point uniformly from the interior of a polytope in R^n . On the other side, due to its generality this inequality is weaker than ours even in a simple case of the sum of n independent exponential random variables ($q = 1$). In this case our concentration inequality gives Chernoff bounds like estimates while Theorem 1.10 provides a much weaker bound. Another drawback of the Theorem 1.10 that it does not handle discrete distributions.

1.7 Paper Outline

We now outline the rest of this paper.

In Section 2 we state and prove several lemmas about the moments of "centered" polynomials that form the heart of our results. In Section 3 we extend these lemmas to moments of arbitrary polynomials. In Section 4 we use these Lemmas to prove our main Theorem 1.4. In Section 5 we prove a counting lemma used in Section 2. We prove our permanent Theorems 1.5 and 1.6 in Section 6. We prove our lower bound Theorem 1.3 in Section 8. We conclude with examples of moment bounded random variables in Section 7.

In Appendix A we prove a special case of our main result: the linear case $q = 1$, i.e. concentration of a sum of independent moment-bounded random variables. This linear case of our theorem is not new, but the proof nicely illustrates many of our techniques with minimal technical complications. The interested reader may find it helpful to study the special cases in Section 1.3 and Appendix A before reading the main body of this paper.

2 Moment Lemma for Centered Polynomials

The proof of the Theorem 1.4 will follow from the application of the Markov's inequality to the upper bound on the k -th moment of the polynomial in question. The first step is to look at moments of “centered” polynomials that replace $Y_v^{\tau_{hv}}$ with $(Y_v^{\tau_{hv}} - \mathbb{E}[Y_v^{\tau_{hv}}])$. For simplicity the heart of our analysis will assume that all coefficients w_h are non-negative; negative coefficients will return in Section 3.

Lemma 2.1 (Initial Moment Lemma) *We are given a hypergraph $H = ([n], \mathcal{H})$, n independent moment bounded random variables $Y = (Y_1, \dots, Y_n)$ with the same parameter L and a polynomial*

$$g(y) = \sum_{h \in \mathcal{H}} w_h \prod_{v \in h} (y_v^{\tau_{hv}} - \mathbb{E}[Y_v^{\tau_{hv}}])$$

with nonnegative coefficients $w_h \geq 0$ such that every monomial (or hyperedge) $h \in \mathcal{H}$ has cardinality exactly η , total power exactly q and maximal power upper bounded by Γ , i.e. $q(h) = q$, $\eta(h) = \eta$ and $\Gamma \geq \max_{v \in h} \tau_{hv}$. It follows that for any integer $k \geq 1$ we have

$$\left| \mathbb{E}[g(Y)^k] \right| \leq \max_{\vec{\nu}} \left\{ R_2^{qk} L^{qk-\Delta} \cdot \Gamma^{qk-\Delta} \cdot k^{qk-(q-1)\nu_0-\Delta} \cdot \left(\prod_{t=0}^q \mu_t^{\nu_t} \right) \right\} \quad (2.8)$$

where $R_2 \geq 1$ is some absolute constant, $\Delta = \sum_{t=0}^q (q-t)\nu_t$, and the maximum is over all non-negative integers ν_t , $0 \leq t \leq q$ satisfying $\nu_0 \leq \nu_q$, $\sum_{t=0}^q \nu_t = k$ and $qk - (q-1)\nu_0 - \Delta \geq 1$ (note also that μ_t are defined according to (1.7), i.e. they depend on original (not centered) random variables Y_v for $v \in [n]$).

Proof. Fix hypergraph $H = ([n], \mathcal{H})$, random variables $Y = (Y_1, \dots, Y_n)$, non-negative weights $\{w_h\}_{h \in \mathcal{H}}$, integer $k \geq 1$, cardinality η and total power q . Without loss of generality we assume that \mathcal{H} is the complete hypergraph (setting additional edge weights to 0 as needed), i.e. \mathcal{H} includes every possible hyperedge over vertex set $[n]$ with cardinality η , total power q , and maximal power at most Γ . A *labeled* hypergraph $G = (\mathcal{V}(G), \mathcal{H}(G))$ consists of a set of vertices $\mathcal{V}(G)$ and a *sequence* of k (not necessarily distinct) hyperedges $\mathcal{H}(G) = h_1, \dots, h_k$. In other words a labeled hypergraph is a hypergraph whose k hyperedges are given unique labels from $[k]$. We write e.g. $\prod_{h \in \mathcal{H}(G)} w_h$ as a shorthand for $\prod_{i=1}^k w_{h_i}$ where $\mathcal{H}(G) = h_1, \dots, h_k$; in particular duplicate hyperedges count multiple times in such a product.

Consider the sequence of hyperedges $h_1, \dots, h_k \in \mathcal{H}$ from our original hypergraph H . These hyperedges define a labeled hypergraph $H(h_1, \dots, h_k)$ with vertex set $\cup_{i=1}^k \mathcal{V}(h_i)$ and hyperedge sequence h_1, \dots, h_k . Note that the vertices of $H(h_1, \dots, h_k)$ are labeled by the indices from $[n]$ and the edges are labeled by the indices from $[k]$. Note also that some hyperedges in $H(h_1, \dots, h_k)$ could span the same set of vertices and have the same power vector, i.e. they are multiple copies of the same hyperedge in the original hypergraph H . Let $\mathcal{P}(H, k)$ be the set of all such edge and vertex labeled hypergraphs that can be generated by any k hyperedges from H . We say that the *degree* of a vertex (in a hypergraph) is the number of hyperedges it appears in. Let $\mathcal{P}_2(H, k) \subseteq \mathcal{P}(H, k)$ be the set of such labeled hypergraphs where each vertex has degree at least two. We split the whole proof into more digestible pieces by subsections.

2.1 Changing the vertex labeling

In this section we will show how to transform the formula for the k -th moment to have the summation over the hypergraphs that have its own set of labels instead of being labeled by the set $[n]$. Let

$X_{hv} = Y_v^{\tau_{hv}} - \mathbb{E}[Y_v^{\tau_{hv}}]$ for $h \in \mathcal{H}$ and $v \in h$. By linearity of expectation, independence of random variables X_{hv} for different vertices $v \in \mathcal{V}$ and definition of $\mathcal{P}(H, k)$ we obtain

$$\begin{aligned}
\left| \mathbb{E} \left[g(Y)^k \right] \right| &= \left| \sum_{h_1, \dots, h_k \in \mathcal{H}} \mathbb{E} \left[\prod_{i=1}^k \left(w_{h_i} \prod_{v \in \mathcal{V}(h_i)} X_{h_i v} \right) \right] \right| \\
&= \left| \sum_{G \in \mathcal{P}(H, k)} \mathbb{E} \left[\prod_{h \in \mathcal{H}(G)} \left(w_h \prod_{v \in \mathcal{V}(h)} X_{hv} \right) \right] \right| \\
&= \left| \sum_{G \in \mathcal{P}(H, k)} \left(\prod_{h \in \mathcal{H}(G)} w_h \right) \left(\prod_{v \in \mathcal{V}(G)} \mathbb{E} \left[\prod_{h \in \mathcal{H}(G) | v \in \mathcal{V}(h)} X_{hv} \right] \right) \right| \\
&= \left| \sum_{G \in \mathcal{P}_2(H, k)} \left(\prod_{h \in \mathcal{H}(G)} w_h \right) \left(\prod_{v \in \mathcal{V}(G)} \mathbb{E} \left[\prod_{h \in \mathcal{H}(G) | v \in \mathcal{V}(h)} X_{hv} \right] \right) \right| \\
&\leq \sum_{G \in \mathcal{P}_2(H, k)} \left(\prod_{h \in \mathcal{H}(G)} w_h \right) \left(\prod_{v \in \mathcal{V}(G)} \left| \mathbb{E} \left[\prod_{h \in \mathcal{H}(G) | v \in \mathcal{V}(h)} X_{hv} \right] \right| \right), \quad (2.9)
\end{aligned}$$

where the last equality follows from the fact that $\mathbb{E}[X_{hv}] = 0$ for all $h \in \mathcal{H}$ and $v \in \mathcal{V}(h)$. Below we will use the notation $\Lambda_v(G) = \left| \mathbb{E} \left[\prod_{h \in \mathcal{H}(G) | v \in \mathcal{V}(h)} X_{hv} \right] \right|$.

Note that a labeled hypergraph $G \in \mathcal{P}_2(H, k)$ could have the number of vertices ranging from η up to $k\eta/2$ since every vertex has degree at least two. For q, η and Γ clear from context, let $\mathcal{S}_2(k, \ell)$ be the set of labeled hypergraphs with vertex set $[\ell]$ having k hyperedges such that each hyperedge has cardinality exactly η , total power q , maximal power $\leq \Gamma$, and every vertex has degree at least 2. For each hypergraph $G \in \mathcal{S}_2(k, \ell)$ the vertices are labeled by the indices from the set $[\ell]$ and the edges are labeled by the indices from the set $[k]$. Let $M(S)$ for $S \subseteq [\ell]$ be the set of all possible injective functions $\pi : S \rightarrow [n]$, in particular $M([\ell])$ is the set of all possible injective functions $\pi : [\ell] \rightarrow [n]$. We will use the notation $\pi(h)$ for a copy of hyperedge $h = (\mathcal{V}(h), \tau(h)) \in \mathcal{H}(G)$ with its vertices relabeled by injective function π , i.e. $\mathcal{V}(\pi(h)) = \{\pi(v) : v \in \mathcal{V}(h)\}$ and $\tau_{\pi(h), \pi(v)} = \tau_{hv}$. Analogously we will use notation $\pi(G)$ to denote the graph G with vertices re-labeled according to function π . We claim that

$$\begin{aligned}
&\sum_{G \in \mathcal{P}_2(H, k)} \left(\prod_{h \in \mathcal{H}(G)} w_h \right) \left(\prod_{v \in \mathcal{V}(G)} \Lambda_v(G) \right) \\
&= \sum_{\ell=\eta}^{k\eta/2} \frac{1}{\ell!} \sum_{G' \in \mathcal{S}_2(k, \ell)} \sum_{\pi \in M([\ell])} \left(\prod_{h \in \mathcal{H}(G')} w_{\pi(h)} \right) \left(\prod_{u \in \mathcal{V}(G')} \Lambda_{\pi(u)}(\pi(G')) \right). \quad (2.10)
\end{aligned}$$

Indeed, every labeled hypergraph $G = (\mathcal{V}(G), \mathcal{H}(G)) \in \mathcal{P}_2(H, k)$ on ℓ vertices has $\ell!$ labeled hypergraphs $G' = (\mathcal{V}(G'), \mathcal{H}(G')) \in \mathcal{S}_2(k, \ell)$ that differ from G by vertex labellings only. Each of those hypergraphs has one corresponding mapping π that maps its ℓ vertex labels into vertex labels of hypergraph $G \in \mathcal{P}_2(H, k)$.

Then, combining (2.9) and (2.10) we obtain

$$\left| \mathbb{E} \left[g(Y)^k \right] \right| \leq \sum_{\ell=\eta}^{k\eta/2} \frac{1}{\ell!} \sum_{G' \in \mathcal{S}_2(k, \ell)} \sum_{\pi \in M([\ell])} \left(\prod_{h \in \mathcal{H}(G')} w_{\pi(h)} \right) \left(\prod_{u \in \mathcal{V}(G')} \Lambda_{\pi(u)}(\pi(G')) \right). \quad (2.11)$$

2.2 Estimating the term for each hypergraph G'

We now fix integer ℓ and labeled hypergraph $G' \in \mathcal{S}_2(k, \ell)$. Let c be the number of *connected components* in G' , i.e. c is a maximal number such that the vertex set $\mathcal{V}(G')$ can be partitioned into c parts $\mathcal{V}_1, \dots, \mathcal{V}_c$ such that for each hyperedge $h \in \mathcal{H}(G')$ and any $j \in [c]$ if $\mathcal{V}(h) \cap \mathcal{V}_j \neq \emptyset$ then $\mathcal{V}(h) \subseteq \mathcal{V}_j$. Intuitively, we can split the vertex set of G' into c components such that there are no hyperedges that have vertices in two or more components. For each vertex $v \in \mathcal{V}(G')$, we define $D_v = \sum_{h \in \mathcal{H}(G') | v \in \mathcal{V}(h)} \tau_{hv}$ to be the sum of all the powers that correspond to the vertex v and hyperedges that are incident to v . We call D_v the *total power* of v . Let d_v denote the number of hyperedges $h \in \mathcal{H}(G')$ with $v \in \mathcal{V}(h)$. We will call d_v the *degree* of the vertex v . By definitions $\sum_{v \in \mathcal{V}(G')} d_v = \eta k$, $\sum_{v \in \mathcal{V}(G')} D_v = qk$ and $d_v \geq 2$ for all $v \in \mathcal{V}(G')$.

We consider a certain canonical ordering $h^{(1)}, \dots, h^{(k)}$ of the hyperedges in $\mathcal{H}(G')$ that will be specified later in Lemma 2.4. (This ordering is distinct from and should not be confused with the ordering of the hyperedges inherent in a labeled hypergraph.) We iteratively remove hyperedges from the hypergraph G' in this order. Let $G'_s = (\mathcal{V}'_s, \mathcal{H}'_s)$ be the hypergraph defined by the hyperedges $\mathcal{H}'_s = h^{(s)}, \dots, h^{(k)}$ and vertex set $\mathcal{V}'_s = \cup_{h \in \mathcal{H}'_s} \mathcal{V}(h)$. In particular G'_1 is identical to G' except for the order of the hyperedges. Let V_s be the vertices of the hyperedge $h^{(s)}$ that have degree one in the hypergraph G'_s , i.e. $\mathcal{V}'_{s+1} = \mathcal{V}'_s \setminus V_s$. By definition, $0 \leq |V_s| \leq \eta$. For each vertex $v \in V_s$, let $\delta_v = \tau_{h^{(s)}v}$, i.e. δ_v is the power corresponding to the vertex v and the last hyperedge in the defined order that is incident to v . We call δ_v the *last power* of v . Intuitively, we delete edges in the order $h^{(1)}, \dots, h^{(k)}$. We also delete all vertices that become isolated. Then V_s is the set of vertices that get deleted during step s of this process.

Recall $[\ell] = \mathcal{V}(G')$ and $\mathcal{V}'_s = \mathcal{V}(G') \setminus \cup_{t=1}^{s-1} V_t$ for $s = 1, \dots, k$. Then

$$\begin{aligned} & \sum_{\pi \in M(\mathcal{V}'_s)} \left(\prod_{h \in \mathcal{H}'_s} w_{\pi(h)} \right) \left(\prod_{v \in \mathcal{V}'_s} \Lambda_{\pi(v)}(\pi(G')) \right) = \\ & \sum_{\pi' \in M(\mathcal{V}'_{s+1})} \sum_{\substack{\pi \in M(\mathcal{V}'_s) \\ \text{s.t. } \pi \text{ extends } \pi'}} \left(\prod_{h \in \mathcal{H}'_{s+1}} w_{\pi(h)} \right) \left(\prod_{v \in \mathcal{V}'_{s+1}} \Lambda_{\pi(v)}(\pi(G')) \right) \left(w_{\pi(h^{(s)})} \prod_{v \in V_s} \Lambda_{\pi(v)}(\pi(G')) \right) = \\ & \sum_{\pi' \in M(\mathcal{V}'_{s+1})} \left[\left(\prod_{h \in \mathcal{H}'_{s+1}} w_{\pi'(h)} \right) \left(\prod_{v \in \mathcal{V}'_{s+1}} \Lambda_{\pi'(v)}(\pi(G')) \right) \sum_{\substack{\pi \in M(\mathcal{V}'_s) \\ \text{s.t. } \pi \text{ extends } \pi'}} \left(w_{\pi(h^{(s)})} \prod_{v \in V_s} \Lambda_{\pi(v)}(\pi(G')) \right) \right] \end{aligned}$$

where we say that π *extends* π' if $\pi(v) = \pi'(v)$ for every v in the domain of π' . By Lemma 2.3 which is an implication of the moment boundness of random variables (Section 2.5) we have

$$\begin{aligned} \Lambda_{\pi(v)}(\pi(G')) & \leq \frac{2^{d_{\pi(v)}} L^{D_{\pi(v)} - \tau_{\pi(h^{(s)})\pi(v)}} \cdot D_{\pi(v)}! \cdot \mathbb{E} \left[\left| Y_{\pi(v)}^{\tau_{\pi(h^{(s)})\pi(v)}} \right| \right]}{\tau_{\pi(h^{(s)})\pi(v)}!} \\ & = \frac{2^{d_v} L^{D_v - \delta_v} \cdot D_v! \cdot \mathbb{E} \left[\left| Y_{\pi(v)}^{\tau_{\pi(h^{(s)})\pi(v)}} \right| \right]}{\delta_v!} \end{aligned}$$

since $\delta_v = \tau_{\pi(h^{(s)})\pi(v)}$ (the vertex degrees and powers do not depend on vertex labeling). Therefore,

$$\sum_{\substack{\pi \in M(\mathcal{V}'_s) \\ \text{s.t. } \pi \text{ extends } \pi'}} \left(w_{\pi(h^{(s)})} \prod_{v \in V_s} \Lambda_{\pi(v)}(\pi(G')) \right) \leq \prod_{v \in V_s} \frac{2^{d_v} L^{D_v - \delta_v} \cdot D_v!}{\delta_v!} \cdot \sum_{\substack{\pi \in M(\mathcal{V}'_s) \\ \text{s.t. } \pi \text{ extends } \pi'}} \left(w_{\pi(h^{(s)})} \prod_{v \in V_s} \mathbb{E} \left[\left| Y_{\pi(v)}^{\tau_{\pi(h^{(s)})\pi(v)}} \right| \right] \right).$$

We now group the sum over π by the value of $\pi(h^{(s)}) \equiv h \in \mathcal{H}$. Note that for any fixed mapping $\pi' \in M(\mathcal{V}'_{s+1})$ there are at most $|V_s|!$ possible mappings $\pi \in M(\mathcal{V}'_s)$ that extend π' and map the vertex labels of hyperedge $h^{(s)} \in G'$ into vertex labels of the hyperedge $h \in \mathcal{H}$. Let $h^{(s)} \setminus V_s$ denote $h^{(s)}$ but with vertices V_s removed, i.e. $\mathcal{V}(h^{(s)} \setminus V_s) = \mathcal{V}(h^{(s)}) \setminus V_s$ and $\tau_{h^{(s)} \setminus V_s, v} = \tau_{h^{(s)} v}$ for all $v \in \mathcal{V}(h^{(s)} \setminus V_s)$. Let $h' = \pi'(h^{(s)} \setminus V_s)$, which is the portion of $\pi(h^{(s)})$ that is fixed by π' . Recall, we write $h \succeq h'$ if $\mathcal{V}(h) \supseteq \mathcal{V}(h')$ and $\tau_{hv} = \tau_{h'v}$ for all $v \in \mathcal{V}(h')$. Also recall the notation $\delta_v = \tau_{h^{(s)} v}$ for $v \in V_s$. Then

$$\begin{aligned} \sum_{\substack{\pi \in M(\mathcal{V}'_s) \\ \text{s.t. } \pi \text{ extends } \pi'}} w_{\pi(h^{(s)})} \prod_{v \in V_s} \mathbb{E} \left[\left| Y_{\pi(v)}^{\tau_{\pi(h^{(s)})\pi(v)}} \right| \right] &\leq |V_s|! \sum_{h \in \mathcal{H} | h \succeq h'} w_h \prod_{u \in \mathcal{V}(h) \setminus \mathcal{V}(h')} \mathbb{E} [|Y_u^{\tau_{hu}}|] \\ &\leq |V_s|! \max_{h' | q(h') = q - \sum_{v \in V_s} \delta_v} \left\{ \sum_{h \in \mathcal{H} | h \succeq h'} w_h \prod_{v \in \mathcal{V}(h) \setminus \mathcal{V}(h')} \mathbb{E} [|Y_v^{\tau_{hv}}|] \right\} \\ &= |V_s|! \mu_{q - \sum_{v \in V_s} \delta_v}. \end{aligned}$$

We repeat the argument for $s = 1, \dots, k$. In the end we obtain

$$\begin{aligned} \sum_{\pi \in M([\ell])} \left(\prod_{h \in \mathcal{H}(\pi(G'))} w_{\pi(h)} \right) \left(\prod_{v \in \mathcal{V}(G')} \Lambda_{\pi(v)}(\pi(G')) \right) &= \\ \sum_{\pi \in M([\ell])} \left(\prod_{h \in \mathcal{H}'_1} w_{\pi(h)} \right) \left(\prod_{v \in \mathcal{V}'_1} \Lambda_{\pi(v)}(G') \right) &\leq \\ \left(\prod_{v \in \mathcal{V}(G')} \frac{2^{d_v} L^{D_v - \delta_v} \cdot D_v!}{\delta_v!} \right) \prod_{s=1}^k (|V_s|! \mu_{q - \sum_{v \in V_s} \delta_v}) &= \\ 2^{\eta k} L^{qk - \Delta} \left(\prod_{v \in \mathcal{V}(G')} \frac{D_v!}{\delta_v!} \right) \left(\prod_{s=1}^k |V_s|! \right) \prod_{t=0}^q \mu_t^{\nu_t} &\leq \\ 2^{\eta k} L^{qk - \Delta} \left(\prod_{v \in \mathcal{V}(G')} \frac{D_v!}{\delta_v!} \right) \eta^\ell \cdot \prod_{t=0}^q \mu_t^{\nu_t} &\quad (2.12) \end{aligned}$$

where ν_t is the number of indices $s = 1, \dots, k$ with $q - \sum_{v \in V_s} \delta_v = t$, $\mu_t = \mu_t(w, Y)$, and $\Delta = \sum_{v \in \mathcal{V}(G')} \delta_v$. In the last inequality we used the fact that $\sum_{s=1}^k |V_s| = \ell$ and $|V_s| \leq \eta$. The quantities ν_t must satisfy the equality $\sum_{t=0}^q (q-t)\nu_t = \sum_{v \in \mathcal{V}(G')} \delta_v = \Delta$.

The flexibility in the choice of the ordering $h^{(1)}, \dots, h^{(k)}$ affects the quality of the bound (2.12) via its influence on the ν_t , δ_v and Δ . We focus on minimizing ν_0 , which intuitively makes sense as μ_0 is often much larger than the other μ_t . The last hyperedge in each of the c connected components must contribute to ν_0 , so $\nu_0 \geq c$. It turns out that equality $\nu_0 = c$ is achievable; the intuition is to pick an ordering that never splits a connected component of G'_s into several components. We defer the proof that such an ordering exists to Lemma 2.4 in Section 2.5. We also know that $\nu_q \geq c$ because the first hyperedge in each connected component is incident to vertices of degree two or more only and therefore contributes to ν_q .

2.3 Using the Counting Lemma

We assume that each hypergraph in $\mathcal{S}_2(k, \ell)$ has an associated canonical ordering of hyperedges, formally defined in Lemma 2.4. This canonical ordering specifies the last powers for all vertices in the hypergraph and the values ν_t for $t = 0, \dots, q$.

We decompose $\mathcal{S}_2(k, \ell)$ as $\mathcal{S}_2(k, \ell) = \bigcup_{c, \bar{d} \geq \bar{2}, \bar{D}, \bar{\delta}} \mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})$ where $\bar{2}$ is a vector of ℓ twos and $\mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})$ is the set of vertex and hyperedge labeled hypergraphs with vertex set $[\ell]$ and k hyperedges such that each hyperedge has cardinality η , total power q , maximal power $\leq \Gamma$, the number of connected components is c , the degree vector is \bar{d} , the total power vector is \bar{D} , and $\bar{\delta}$ is the vector of last powers (corresponding to the canonical ordering). Note that $\mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})$ depends on q , η and Γ as well. Let $\bar{\nu} = (\nu_0, \dots, \nu_q)$. Combining, (2.11) and (2.12) we obtain

$$\begin{aligned} \left| \mathbb{E} \left[g(Y)^k \right] \right| &\leq \sum_{\ell=\eta}^{k\eta/2} \frac{1}{\ell!} \sum_{c=1}^{\ell/\eta} \sum_{\bar{d} \geq \bar{2}, \bar{D}, \bar{\delta}} \sum_{G' \in \mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})} 2^{\eta k} L^{qk-\Delta} \left(\prod_{t=0}^q \mu_t^{\nu_t} \right) \eta^\ell \prod_{v \in [\ell]} \frac{D_v!}{\delta_v!} \\ &\leq \max_{\ell, c, \bar{d} \geq \bar{2}, \bar{D}, \bar{\delta}, \bar{\nu}} \left\{ \frac{k\eta}{2} \cdot \frac{1}{\ell!} \cdot \frac{\ell}{\eta} \cdot (2^{qk+\ell})^3 \cdot |\mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})| \cdot 2^{\eta k} L^{qk-\Delta} \left(\prod_{t=0}^q \mu_t^{\nu_t} \right) \eta^\ell \prod_v \frac{D_v!}{\delta_v!} \right\} \\ &\leq \max_{\ell, c, \bar{d} \geq \bar{2}, \bar{D}, \bar{\delta}, \bar{\nu}} \left\{ \frac{k\ell}{2 \cdot \ell!} \cdot 2^{3(qk+\ell)} \cdot 2^{\eta k} L^{qk-\Delta} \cdot \eta^\ell \cdot \underbrace{R_0^{qk} \cdot \Gamma^{qk-\Delta-\ell} \cdot k^{qk-(q-1)c-\Delta+\ell}}_{\substack{|\mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})| \prod_v \frac{D_v!}{\delta_v!} \leq \text{this} \\ \text{by counting Lemma 5.1 (Section 5)}}} \cdot \left(\prod_{t=0}^q \mu_t^{\nu_t} \right) \right\} \end{aligned}$$

where the maximum over $\bar{\nu}$ is over $\nu_0, \dots, \nu_q \geq 0$ with $c = \nu_0 \leq \nu_q$, $\sum_{t=0}^q \nu_t = k$ and $\sum_{t=0}^q (q-t)\nu_t = \Delta$. Also the integers ν_0, \dots, ν_q must satisfy the inequality $(q-1)k - \Delta \geq (q-2)\nu_0$ by Corollary 2.6. The second inequality follows from the fact that the total number of feasible total power vectors \bar{D} is at most $2^{qk+\ell}$ (qk is the sum of all the powers and we need to compute the total number of partitions of the array with qk entries into ℓ possible groups of consecutive entries which is $\binom{qk+\ell-1}{\ell-1}$), and similarly the number of vectors \bar{d} and $\bar{\delta}$ can also be upper-bounded by $2^{qk+\ell}$. We substitute ν_0 for c , and remove the unreferenced variables c , \bar{d} , \bar{D} and $\bar{\delta}$ from the maximum. The maximum over $\bar{\nu}$ is now over $\nu_0, \dots, \nu_q \geq 0$ with $\nu_0 \leq \nu_q$, $\sum_{t=0}^q \nu_t = k$ such that $(q-1)k - \Delta \geq (q-2)\nu_0$. We

continue

$$\begin{aligned}
\left| \mathbb{E} \left[g(Y)^k \right] \right| &\leq \max_{\ell, \bar{\nu}} \left\{ \frac{k\ell}{2 \cdot \ell!} \cdot 2^{3(qk+\ell)} \cdot 2^{\eta k} L^{qk-\Delta} \cdot \eta^\ell \cdot R_0^{qk} \cdot \Gamma^{qk-\Delta-\ell} \cdot k^{qk-(q-1)\nu_0-\Delta+\ell} \cdot \left(\prod_{t=0}^q \mu_t^{\nu_t} \right) \right\} \\
&\leq \max_{\ell, \bar{\nu}} \left\{ R_1^{qk} \cdot \left(\frac{\eta}{\ell} \right)^\ell \cdot L^{qk-\Delta} \cdot \Gamma^{qk-\Delta} \cdot k^{qk-(q-1)\nu_0-\Delta+\ell} \cdot \left(\prod_{t=0}^q \mu_t^{\nu_t} \right) \right\} \\
&= \max_{\ell, \bar{\nu}} \left\{ R_1^{qk} \cdot \left(\frac{\eta k}{\ell} \right)^\ell \cdot L^{qk-\Delta} \cdot \Gamma^{qk-\Delta} \cdot k^{qk-(q-1)\nu_0-\Delta} \cdot \left(\prod_{t=0}^q \mu_t^{\nu_t} \right) \right\} \\
&\leq \max_{\bar{\nu}} \left\{ R_2^{qk} L^{qk-\Delta} \cdot \Gamma^{qk-\Delta} \cdot k^{qk-(q-1)\nu_0-\Delta} \cdot \left(\prod_{t=0}^q \mu_t^{\nu_t} \right) \right\} \tag{2.13}
\end{aligned}$$

where $R_0 < R_1 < R_2$ are some absolute constants, the second inequality uses the facts that $\ell! \geq (\ell/e)^\ell$ and $\Gamma^{-\ell} \leq 1$, and the last inequality is implied by the fact that

$$\left(\frac{k\eta}{\ell} \right)^\ell \leq \max_{x>0} \left(\frac{k\eta}{x} \right)^x = e^{k\eta/e}.$$

Inequality (2.13) is precisely the inequality (2.8) that we needed to prove. Note that the inequality $(q-1)k - \Delta \geq (q-2)\nu_0$ implies $qk - (q-1)\nu_0 - \Delta \geq k - \nu_0 = k - c \geq c \geq 1$ (we use the fact that each connected component has at least two hyperedges). \blacksquare

2.4 Intermediate moment lemma

Lemma 2.2 (Intermediate Moment Lemma) *We are given n independent moment bounded random variables $Y = (Y_1, \dots, Y_n)$ with the same parameter L and a general polynomial $f(x)$ with nonnegative coefficients such that every monomial (or hyperedge) $h \in \mathcal{H}$ has exactly η variables, total power exactly q and power of any variable upper bounded by Γ , i.e. $q(h) = q$, $\eta(h) = \eta$ and $\Gamma = \max_{h \in \mathcal{H}, v \in h} \tau_{hv}$. Then*

$$\left| \mathbb{E} \left[g(Y)^k \right] \right| \leq \max \left\{ \left(\sqrt{k R_3^q \Gamma^q L^q \mu_q \mu_0} \right)^k, \max_{t \in [q]} (k^t R_3^q L^t \Gamma^t \mu_t)^k \right\}. \tag{2.14}$$

where $R_3 \geq 1$ is some absolute constant, $g(Y)$ is a polynomial of centered random variables defined in Lemma 2.1 and $[q] = \{1, \dots, q\}$.

Proof. We apply Lemma 2.1. Since $\sum_{t=0}^q \nu_t = k$ and $\sum_{t=0}^q (q-t)\nu_t = \Delta$ we have,

$$\begin{aligned}
q(\nu_q - \nu_0) + \nu_0 + \sum_{t=1}^{q-1} t\nu_t &= -(q-1)\nu_0 + \sum_{t=0}^q t\nu_t \\
&= q \left(k - \sum_{t=0}^q \nu_t \right) - (q-1)\nu_0 + \sum_{t=0}^q t\nu_t \\
&= qk - (q-1)\nu_0 - \sum_{t=0}^q (q-t)\nu_t \\
&= qk - (q-1)\nu_0 - \Delta, \\
q(\nu_q - \nu_0) + q\nu_0 + \sum_{t=1}^{q-1} t\nu_t &= qk - \Delta.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \max_{\bar{\nu}} \left\{ R_2^{qk} L^{qk-\Delta} \cdot \Gamma^{qk-\Delta} \cdot k^{qk-(q-1)\nu_0-\Delta} \cdot \left(\prod_{t=0}^q \mu_t^{\nu_t} \right) \right\} \\
&= \max_{\bar{\nu}} \left\{ (k^q R_2^q \Gamma^q L^q \mu_q)^{\nu_q-\nu_0} \cdot (k R_2^{2q} \Gamma^q L^q \mu_q \mu_0)^{\nu_0} \cdot \prod_{t=1}^{q-1} (k^t R_2^q \Gamma^t L^t \mu_t)^{\nu_t} \right\} \\
&\leq \max_{\bar{\nu}} \left\{ (k^q R_3^q \Gamma^q L^q \mu_q)^{\nu_q-\nu_0} \cdot (k R_3^q \Gamma^q L^q \mu_q \mu_0)^{\nu_0} \cdot \prod_{t=1}^{q-1} (k^t R_3^q \Gamma^t L^t \mu_t)^{\nu_t} \right\}
\end{aligned}$$

for the absolute constant $R_3 = R_2^2$. Using the facts that $\sum_{t=0}^q \nu_t = k$ and $\nu_q \geq \nu_0$ again, we derive,

$$\begin{aligned}
\left| \mathbb{E} [g(Y)^k] \right| &\leq \max_{\bar{\nu}} \left\{ (k^q R_3^q \Gamma^q L^q \mu_q)^{\nu_q-\nu_0} \cdot \left(\sqrt{k R_3^q \Gamma^q L^q \mu_q \mu_0} \right)^{2\nu_0} \cdot \prod_{t=1}^{q-1} (k^t R_3^q \Gamma^t L^t \mu_t)^{\nu_t} \right\} \\
&\leq \max \left\{ (k^q R_3^q \Gamma^q L^q \mu_q)^k, \left(\sqrt{k R_3^q \Gamma^q L^q \mu_q \mu_0} \right)^k, \max_{t \in [q-1]} (k^t R_3^q \Gamma^t L^t \mu_t)^k \right\} \\
&= \max \left\{ \left(\sqrt{k R_3^q \Gamma^q L^q \mu_q \mu_0} \right)^k, \max_{t \in [q]} (k^t R_3^q \Gamma^t L^t \mu_t)^k \right\}.
\end{aligned}$$

■

2.5 Three Technical Lemmas

In this section we prove technical lemmas that were used in the proof of Lemma 2.1.

Lemma 2.3 *For any moment bounded random variable Z with parameter L , integer $k \geq 1$, set $S \subseteq [k]$ and a collection of positive integer powers d_t for $t \in S$, the following inequality holds:*

$$\left| \mathbb{E} \left[\prod_{t \in S} (Z^t)^{d_t} \right] \right| \leq \min_{t \in S} \left\{ \frac{L^{D-t} \cdot D! \cdot \mathbb{E} [|Z|^t]}{t!} \right\}$$

and

$$\left| \mathbb{E} \left[\prod_{t \in S} (Z^t - \mathbb{E} [Z^t])^{d_t} \right] \right| \leq \min_{t \in S} \left\{ \frac{2^d L^{D-t} \cdot D! \cdot \mathbb{E} [|Z|^t]}{t!} \right\}$$

where $D = \sum_{t \in S} t d_t$ and $d = \sum_{t \in S} d_t$.

Proof. To prove the first inequality note that for any $\tau \in S$ by Jensen's inequality we have

$$\begin{aligned}
\left| \mathbb{E} \left[\prod_{t \in S} (Z^t)^{d_t} \right] \right| &= |\mathbb{E} [Z^D]| \leq \mathbb{E} [|Z|^D] \\
&\leq \frac{L^{D-\tau} D!}{\tau!} \mathbb{E} [|Z|^\tau]
\end{aligned} \tag{2.15}$$

where the final inequality follows from applying Definition 1.1 $D - \tau$ times.

To show the second inequality we bound

$$\begin{aligned}
\left| \mathbb{E} \left[\prod_{t \in S} (Z^t - \mathbb{E}[Z^t])^{d_t} \right] \right| &\leq \prod_{t \in S} \left(\mathbb{E} \left[|Z^t - \mathbb{E}[Z^t]|^{d_t \cdot \frac{D}{d_t}} \right] \right)^{td_t/D} \\
&\leq \prod_{t \in S} \left(\mathbb{E} \left[(|Z^t| + \mathbb{E}[|Z^t|])^{\frac{D}{d_t}} \right] \right)^{td_t/D} \\
&\leq \prod_{t \in S} \left(\mathbb{E} \left[2^{\frac{D}{d_t}-1} \left(|Z|^D + (\mathbb{E}[|Z^t|])^{D/d_t} \right) \right] \right)^{td_t/D} \\
&\leq \prod_{t \in S} \left(2^{\frac{D}{d_t}} \mathbb{E}[|Z|^D] \right)^{td_t/D} \\
&= 2^d \mathbb{E}[|Z|^D] \\
&\leq \frac{2^d L^{D-\tau} D!}{\tau!} \mathbb{E}[|Z|^\tau]
\end{aligned}$$

where the first inequality uses Hölder's Inequality (see Lemma 3.1), the third inequality uses the fact (which follows from convexity) that $(x + y)^p \leq 2^{p-1}(x^p + y^p)$ for any $p \geq 1$ and $x, y \geq 0$ (in particular $(x = |Z^t|$ and $y = \mathbb{E}[|Z^t|])$), the fourth inequality uses Jensen's inequality, and the last inequality uses the inequality (2.15). ■

We now prove the following intuitive fact that was left unproven near the end of Section 2.2.

Lemma 2.4 *In the notation of Section 2.2 there exists a canonical ordering $h^{(1)}, \dots, h^{(k)}$ of the hyperedges $\mathcal{H}(G')$ such that $\nu_0 = c$.*

Proof. Let \mathcal{L} be the line graph of G' , i.e. an undirected graph with one vertex for each of the k hyperedges of G' and an edge connecting every pair of vertices that correspond to hyperedges with intersecting vertex sets. We define the desired sequence of hyperedges $h^{(1)}, \dots, h^{(k)}$ and a sequence of induced subgraphs $\mathcal{L}_1, \dots, \mathcal{L}_k$ of \mathcal{L} as follows.

We set \mathcal{L}_1 to \mathcal{L} . For any $1 \leq s \leq k$ we form \mathcal{L}_{s+1} from \mathcal{L}_s by removing vertex $h^{(s)}$, where $h^{(s)}$ has the lowest label from the vertices of \mathcal{L}_s subject to the constraint that the number of connected components n_{s+1} of \mathcal{L}_{s+1} must not exceed the number of connected components n_s of \mathcal{L}_s . For example pick $h^{(s)}$ to be an arbitrary leaf of a depth first search tree started from an arbitrary vertex of \mathcal{L}_s or an isolated vertex of \mathcal{L}_s if there are any.

It remains to show that the ordering $h^{(1)}, \dots, h^{(k)}$ satisfies the desired property $\nu_0 = c$. Note that $h^{(s)}$ contributes to ν_0 if and only if $V_s = \mathcal{V}(h^{(s)})$, that is if and only if $h^{(s)}$ is an isolated vertex in \mathcal{L}_s . Whenever such an $h^{(s)}$ is chosen the number of connected components decreases by one (i.e. $n_{s+1} = n_s - 1$), and otherwise the number of connected components is unchanged (i.e. $n_{s+1} = n_s$). We conclude that $\nu_0 = n_1 - n_{k+1} = c - 0$ as desired. ■

The following Lemma was used in the proof of the Initial Moment Lemma and will be used later in the proof of the Main Counting Lemma.

Lemma 2.5 *Let G' be a labeled hypergraph with all degrees at least two, c connected components with sets of vertices C_1, \dots, C_c and the number of hyperedges k_1, \dots, k_c . Further, let $h^{(1)}, \dots, h^{(k)}$ be the canonical ordering of its hyperedges specified in Lemma 2.4 where $k = \sum_{i=1}^c k_i$. Then for each $i = 1, \dots, c$ we have*

$$(q-1)k_i - \sum_{v \in C_i} \delta_v \geq q-2$$

where δ_v is the last power of vertex v corresponding to the canonical ordering of hyperedges as defined in Section 2.2.

Proof. Fix a labeled hypergraph G' . Recall that the power of a vertex v in hyperedge h is denoted τ_{hv} . Following Section 2.2 define V_j to be the set of vertices whose last incident hyperedge is $h^{(j)}$, where “last” is relative to the canonical ordering $h^{(1)}, \dots, h^{(k)}$ of the hyperedges defined in that section. Let $I_i = \{s \in [k] \mid \mathcal{V}(h^{(s)}) \subseteq C_i\}$. We charge $(q-1)k_i - \sum_{v \in C_i} \delta_v$ to the various hyperedges in the following natural way:

$$(q-1)k_i - \sum_{v \in C_i} \delta_v = \sum_{j \in I_i} \left((q-1) - \sum_{v \in V_j} \tau_{h^{(j)}v} \right) \equiv \sum_{j \in I_i} \alpha_j. \quad (2.16)$$

The contribution α_s of the hyperedge with smallest index $s \in I_i$ is exactly $q-1$ since the degree of each vertex v is at least two and $h^{(s)}$ is the first hyperedge in this connected component that we delete, i.e. $V_s = \emptyset$. The last hyperedge $h^{(s')}$ for $s' = \max\{j \mid j \in I_i\}$ clearly contributes $\alpha_{s'} = (q-1) - q = -1$ since $|V_{s'}| = \eta$. For any $j \in I_i \setminus \{s'\}$ we know that $\sum_{v \in V_j} \tau_{h^{(j)}v} \leq q-1$ because $|V_j| \leq \eta-1$. Otherwise component i would contribute more than 1 to ν_0 and ν_0 would exceed c . We conclude that $\alpha_j \geq 0$ for $j \in I_i \setminus \{s'\}$. Using (2.16) and these lower bounds on the α_j we bound

$$(q-1)k_i - \sum_{v \in C_i} \delta_v = \sum_{j \in I_i} \alpha_j \geq q-1 + 0 - 1 = q-2$$

as desired. ■

The following Corollary immediately follows from Lemma 2.5.

Corollary 2.6 *Let G' be a labeled hypergraph with all degrees at least two, c connected components and the canonical ordering satisfying the condition $\nu_0 = c$ we have $qk - \Delta \geq (q-2)\nu_0$ where $\Delta = \sum_{v \in \mathcal{V}(G)} \delta_v = \sum_{t=0}^q (q-t)\nu_t$.*

3 General Even Moment Lemma

Lemma 3.1 (Hölder’s Inequality) *Let $p_1, \dots, p_k \in (1, +\infty)$ such that $\sum_{i=1}^k \frac{1}{p_i} = 1$ then for arbitrary collection X_1, \dots, X_k of random variables on the same probability space the following inequality holds*

$$\mathbb{E} \left[\left| \prod_{i=1}^k X_i \right| \right] \leq \prod_{i=1}^k \mathbb{E} [|X_i|^{p_i}]^{1/p_i}.$$

We will use the following corollary of Hölder’s inequality known as Minkowski inequality (or triangle inequality for norms).

Corollary 3.2 (Minkowski Inequality) *Let $k \geq 1$ and Z_1, Z_2, \dots, Z_m be (potentially dependent) random variables with $\mathbb{E}[|Z_i|^k] \leq z_i^k$ for $z_i \in \mathbb{R}_+$. It follows that*

$$\mathbb{E} \left[\left(\sum_{i=1}^m |Z_i| \right)^k \right] \leq \left(\sum_{i=1}^m z_i \right)^k. \quad (3.17)$$

Lemma 3.3 (General Even Moment Lemma) *We are given n independent moment bounded random variables $Y = (Y_1, \dots, Y_n)$ with the same parameter L and a general power q polynomial $f(x)$ and maximal variable power $\Gamma = \max_{h \in \mathcal{H}, v \in h} \tau_{hv}$. Let $k \geq 2$ be an even integer then*

$$\mathbb{E} \left[|f(Y) - \mathbb{E}[f(Y)]|^k \right] \leq \max \left\{ \max_{t \in [q]} \left(\sqrt{k R_4^q \Gamma^t L^t \mu_t \mu_0} \right)^k, \max_{t \in [q]} (k^t R_4^q \Gamma^t L^t \mu_t)^k \right\}. \quad (3.18)$$

where $R_4 \geq 1$ is some absolute constant.

Proof. Let weight function w and hypergraph $H = ([n], \mathcal{H})$ be such that $f(Y) = \sum_{h \in \mathcal{H}} w_h \prod_{v \in \mathcal{V}(h)} Y_v^{\tau_{hv}}$. Let $X_{hv} = Y_v^{\tau_{hv}} - \mathbb{E}[Y_v^{\tau_{hv}}]$. Let \mathcal{H}' denote the set of all possible hyperedges (including the empty hyperedge) with vertices from $\mathcal{V}(H) = [n]$ and total power at most q . First we note that

$$\begin{aligned} f(Y) &= \sum_{h \in \mathcal{H}} w_h \prod_{v \in \mathcal{V}(h)} (X_{hv} + \mathbb{E}[Y_v^{\tau_{hv}}]) \\ &= \sum_{h' \in \mathcal{H}'} \sum_{h \in \mathcal{H}: h \supseteq h'} w_h \left(\prod_{v \in \mathcal{V}(h) \setminus \mathcal{V}(h')} \mathbb{E}[Y_v^{\tau_{hv}}] \right) \left(\prod_{v \in \mathcal{V}(h')} X_{hv} \right) \\ &= \sum_{h' \in \mathcal{H}'} w'_{h'} \prod_{v \in \mathcal{V}(h')} X_{h'v} \end{aligned} \quad (3.19)$$

where h' ranges over all possible hyperedges (including the empty hyperedge) and

$$w'_{h'} = \sum_{h \in \mathcal{H}: h \supseteq h'} w_h \left(\prod_{v \in \mathcal{V}(h) \setminus \mathcal{V}(h')} \mathbb{E}[Y_v^{\tau_{hv}}] \right).$$

We next group the monomials on the right hand side of (3.19) by cardinality, power, and sign of coefficient, yielding $m \leq 2q^2$ polynomials $g^{(1)}, \dots, g^{(m)}$ with corresponding weight functions for all monomials $w^{(1)}, \dots, w^{(m)}$ and powers q_1, \dots, q_m . That is,

$$\begin{aligned} f(Y) &= w'_{\{\}} + \sum_{i=1}^m \sum_{h': \eta(h') \geq 1} w_{h'}^{(i)} \prod_{v \in \mathcal{V}(h')} X_{h'v} \\ &= \mathbb{E}[f(Y)] + \sum_{i=1}^m g^{(i)}(Y) \end{aligned} \quad (3.20)$$

where $\{\}$ is the empty hyperedge. We have

$$\begin{aligned} \mu_r(w^{(i)}, Y) &\leq \mu_r(w', Y) = \max_{h_0: q(h_0)=r} \sum_{h' \supseteq h_0} |w'_{h'}| \prod_{v \in \mathcal{V}(h') \setminus \mathcal{V}(h_0)} \mathbb{E}[|Y_v^{\tau_{h'v}}|] \\ &\leq \max_{h_0: q(h_0)=r} \sum_{h' \supseteq h_0} \sum_{h \supseteq h'} |w_h| \left(\prod_{v \in \mathcal{V}(h) \setminus \mathcal{V}(h')} \mathbb{E}[|Y_v^{\tau_{hv}}|] \right) \prod_{v \in \mathcal{V}(h') \setminus \mathcal{V}(h_0)} \mathbb{E}[|Y_v^{\tau_{hv}}|] \\ &\leq \max_{h_0: q(h_0)=r} \sum_{h' \supseteq h_0} \sum_{h \supseteq h'} |w_h| \left(\prod_{v \in \mathcal{V}(h) \setminus \mathcal{V}(h_0)} \mathbb{E}[|Y_v|^{\tau_{hv}}] \right) \\ &\leq 2^q \max_{h_0: q(h_0)=r} \sum_{h \supseteq h_0} |w_h| \left(\prod_{v \in \mathcal{V}(h) \setminus \mathcal{V}(h_0)} \mathbb{E}[|Y_v|^{\tau_{hv}}] \right) = 2^q \mu_r(w, Y) = 2^q \mu_r \end{aligned}$$

where we upper bounded the number of hyperedges h' such that $h \succeq h' \succeq h_0$ by 2^q . Therefore, for even $k \geq 2$ the Lemma 2.2 implies that

$$\mathbb{E} \left[\left| g^{(i)}(Y) \right|^k \right] = \left| \mathbb{E} \left[g^{(i)}(Y)^k \right] \right| \leq 2^{qk} \max \left\{ \left(\sqrt{k R_3^q \Gamma^{q_i} L^{q_i} \mu_{q_i} \mu_0} \right)^k, \max_{t \in [q_i]} (k^t R_3^{q_i} L^t \Gamma^t \mu_t)^k \right\} = 2^{qk} z_i^k.$$

Applying Corollary 3.2 yields

$$\begin{aligned} \mathbb{E} \left[|f(Y) - \mathbb{E}[f(Y)]|^k \right] &\leq \mathbb{E} \left[\left(\sum_{i=1}^m |g_i(Y)| \right)^k \right] \leq 2^{qk} \left(\sum_{i=1}^m z_i \right)^k \leq 2^{qk} m^k \max_{i \in [m]} z_i^k \\ &\leq 2^{qk} m^k \cdot \max \left\{ \max_{t \in [q]} \left(\sqrt{k R_3^q \Gamma^t L^t \mu_t \mu_0} \right)^k, \max_{t \in [q]} (k^t R_3^q L^t \Gamma^t \mu_t)^k \right\} \\ &\leq \max \left\{ \max_{t \in [q]} \left(\sqrt{k R_4^q \Gamma^t L^t \mu_t \mu_0} \right)^k, \max_{t \in [q]} (k^t R_4^q L^t \Gamma^t \mu_t)^k \right\} \end{aligned}$$

for some absolute constant R_4 such that $m^2 2^{2q} R_3^q \leq R_4^q$. ■

4 Proof of the Theorem 1.4

Now we prove Theorem 1.4 by applying the Markov's inequality.

Proof. By Markov's inequality we derive

$$Pr[|f(Y) - \mathbb{E}[f(Y)]| \geq \lambda] = Pr[|f(Y) - \mathbb{E}[f(Y)]|^k \geq \lambda^k] \leq \frac{\mathbb{E}[|f(Y) - \mathbb{E}[f(Y)]|^k]}{\lambda^k}.$$

Choosing $k^* \geq 0$ to be the even integer such that $k^* \in (K - 2, K]$ for

$$K = \min \left\{ \min_{t \in [q]} \frac{\lambda^2}{e^2 R_4^q \Gamma^t L^t \mu_t \mu_0}, \min_{t \in [q]} \left(\frac{\lambda}{e R_4^q L^t \Gamma^t \mu_t} \right)^{1/t} \right\}$$

i.e.

$$\frac{\sqrt{k^* R_4^q \Gamma^t L^t \mu_t \mu_0}}{\lambda} \leq 1/e \text{ and } \frac{(k^*)^t R_4^q L^t \Gamma^t \mu_t}{\lambda} \leq 1/e$$

for all $t \in [q]$. Using the inequality (3.18) from the Lemma 3.3 we derive

$$\begin{aligned} Pr[|f(Y) - \mathbb{E}[f(Y)]| \geq \lambda] &\leq \frac{\mathbb{E}[|f(Y) - \mathbb{E}[f(Y)]|^{k^*}]}{\lambda^{k^*}} \\ &\leq \max \left\{ \max_{t \in [q]} e^{k^* \ln \frac{\sqrt{k^* R_4^q \Gamma^t L^t \mu_t \mu_0}}{\lambda}}, \max_{t \in [q]} e^{k^* \ln \frac{(k^*)^t R_4^q L^t \Gamma^t \mu_t}{\lambda}} \right\} \\ &\leq e^{-k^*} \leq e^{-K+2} \\ &\leq e^2 \cdot \max \left\{ \max_{t \in [q]} e^{-\frac{\lambda^2}{R^q \Gamma^t L^t \mu_t \mu_0}}, \max_{t \in [q]} e^{-\frac{1}{L \cdot \Gamma} \left(\frac{\lambda}{R^q \mu_t} \right)^{1/t}} \right\}, \end{aligned}$$

for some universal constant $R > R_4 \geq 1$. This implies the statement of the Theorem. ■

5 Counting Lemma

In this section we consider labeled hypergraphs in $\mathcal{S}_2(k, \ell)$ for fixed parameters η, q and Γ . We use C_1, \dots, C_c to denote the set of vertices in the connected components of a labeled hypergraph. We use ℓ_1, \dots, ℓ_c and k_1, \dots, k_c to denote the number of vertices and hyperedges in those connected components. We will freely use the following elementary facts:

1. $\eta \leq \ell_i \leq \eta k_i/2$ where the lower bound follows from the fact that each connected component has at least one hyperedge and the upper bound follows from the fact that each vertex has degree at least two;
2. $\eta c \leq \ell \leq \eta k/2$ (these inequalities are obtained by summing up the above inequalities over all connected components);
3. $1 \leq c \leq k/2$, the lower bound is obvious and the upper bound follows from the previous inequality.

In two of the auxiliary lemmas below we will use the classical Gibbs inequality which states that for two arbitrary discrete probability distributions p_1, \dots, p_n and q_1, \dots, q_n with strictly positive p_i, q_i the following inequality holds

$$-\sum_{i=1}^n p_i \log_2 p_i \leq -\sum_{i=1}^n p_i \log_2 q_i$$

or equivalently

$$\prod_{i=1}^n p_i^{p_i} \geq \prod_{i=1}^n q_i^{p_i}.$$

In what follows we identify a vertex with its index $v \in [\ell]$. The main statement of this section is the following

Lemma 5.1 (Main Counting Lemma) *For any $k, \Gamma, q \geq \eta \geq 1, \ell, c, \bar{D}, \bar{d} \geq \bar{2}$ and $\bar{\delta}$ we have*

$$|\mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})| \left(\prod_{v \in [\ell]} \frac{D_v!}{\delta_v!} \right) \leq R_0^{qk} \Gamma^{qk - \ell - \sum_{v \in [\ell]} \delta_v} k^{qk - c(q-1) - \sum_{v \in [\ell]} (\delta_v - 1)},$$

for some universal constant $R_0 > 1$.

We prove Lemma 5.1 as a sequence of auxiliary Lemmas.

We say that C_1, \dots, C_c and k_1, \dots, k_c are *feasible* (with respect to $\bar{d}, \bar{D}, \bar{\delta}$ clear from context) if there is a labeled hypergraph in $\mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})$ with corresponding canonical ordering of its hyperedges whose connected components (numbered arbitrarily) have vertex sets C_1, \dots, C_c and number of hyperedges k_1, \dots, k_c .

Lemma 5.2 *For any k, ℓ, q, η , and \bar{d} we have*

$$|\cup_{c, \bar{D}, \bar{\delta}} \mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})| \leq \binom{q-1}{\eta-1}^k \prod_{v \in [\ell]} \binom{k}{d_v} \leq 2^{(q-1)k} \frac{k^{\eta k}}{\prod_v d_v!}.$$

Note, that we intentionally do not fix Γ since the bound in the Lemma holds for any $\Gamma \leq q$.

Proof. Fix k, ℓ, η, q , and \bar{d} . To show the first inequality, note that a labeled hypergraph is uniquely specified by:

1. for every vertex $v = 1, \dots, \ell$ whether or not it appears in each of the k hyperedges and
2. for every hyperedge h the power vector of its η vertices.

Vertex v with degree d_v clearly has $\binom{k}{d_v}$ possible sets of hyperedges it can appear in, so there are at most $\prod_{v \in [\ell]} \binom{k}{d_v}$ ways to assign vertices to the hyperedges. In general this is quite a rough estimate (since we do not use the fact that each hyperedge contains exactly η vertices). More precisely we would like to estimate the number of 0/1 matrices of dimension $k \times \ell$ with prescribed row sums equal to η and a column sum d_v for a row indexed by v . Estimating this quantity is an important topic in combinatorics (see survey [9] and references therein) but for our purposes the simple estimate above provides a tight bound.

We now count the ways to assign weights to the hyperedges. Recall that τ_{hv} denote the weight of the v^{th} vertex in hyperedge h . There is a standard bijection between $q-1$ digit binary strings with $\eta-1$ zeros and placements of q identical items into η bins such that each bin has at least one item. The string starts with $\tau_{1,h}-1$ ones followed by a zero, followed by $\tau_{2,h}-1$ ones followed by a zero and so on, ending with $\tau_{q,h}-1$ ones (and no trailing zero). We conclude that there are $\binom{q-1}{\eta-1}^k$ ways to assign weights of the hyperedges. This concludes the proof of the first inequality.

The second inequality follows because $\binom{q-1}{\eta-1} \leq 2^{q-1}$, $\binom{k}{d_v} \leq k^{d_v}/d_v!$, and $\sum_v d_v = \eta k$. \blacksquare

Lemma 5.3 *For any $k, \ell, c, q, \eta, \bar{d} \geq \bar{2}, \bar{D}$ and $\bar{\delta}$ we have*

$$|\mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})| \leq 2^{\ell+k+5c+qk} k! \left(\prod_{v \in [\ell]} \frac{1}{d_v!} \right) \max_{\substack{C_1, \dots, C_c \\ k_1, \dots, k_c \\ \text{feasible}}} \left(\prod_{i=1}^c \frac{k_i^{\eta k_i}}{k_i!} \binom{\ell}{|C_i| - 1} \right)$$

where the maximums are evaluated over all C_1, \dots, C_c and k_1, \dots, k_c that are feasible as defined above. This bound holds for any $\Gamma \leq q$.

Proof. We prove the Lemma by mapping the labeled hypergraphs in $\mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})$ into distinct binary strings and bounding the length of these strings.

Fix an arbitrary hypergraph in $\mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})$. Our encoding begins by encoding the vertices in the connected component that contains vertex 1. Let the vertices in this component be denoted by C_1 and $|C_1| = \ell_1$. We encode ℓ_1 in unary, e.g. our string begins with 1110 if $\ell_1 = 3$. We then encode the identity of the remaining $\ell_1 - 1$ vertices in $C_1 \setminus \{1\}$ using a single character with $\left\lceil \log_2 \binom{\ell}{\ell_1 - 1} \right\rceil$ binary digits, i.e. we have $\binom{\ell}{\ell_1 - 1}$ options which are encoded in binary.

We then look at the lowest-indexed vertex that has yet to be placed in a connected component and encode the size ℓ_2 and vertices C_2 of its component in the same manner. We repeat until all ℓ vertices have been placed in one of the c connected components, where the i^{th} component considered has ℓ_i vertices. At this point we have partitioned the vertices into connected components using

$$\sum_{i=1}^c \left[\ell_i + 1 + \left\lceil \log_2 \binom{\ell}{\ell_i - 1} \right\rceil \right] = \ell + c + \sum_{i=1}^c \left\lceil \log_2 \binom{\ell}{\ell_i - 1} \right\rceil \quad (5.21)$$

bits.

We then encode the number of hyperedges k_i in each connected component in unary using

$$\sum_{i=1}^c (k_i + 1) = k + c \quad (5.22)$$

bits, i.e. we have k_1 ones followed by a zero, followed by k_2 ones followed by a zero and so on. There are $\frac{k!}{k_1! \dots k_c!}$ ways to partition k indices into c groups where i -th group has k_i indices. Therefore, we can encode which component each hyperedge is in using

$$\left\lceil \log_2 \left(\frac{k!}{k_1! \dots k_c!} \right) \right\rceil \quad (5.23)$$

bits.

Finally we encode the vertices and the power vectors of each hyperedge in component $1 \leq i \leq c$. The number of possibilities is clearly $|\mathcal{S}(k_i, \ell_i, 1, \bar{d}|_i, \bar{D}|_i, \bar{\delta}|_i)|$, where $\bar{d}|_i$ (resp. $\bar{D}|_i$ and $\bar{\delta}|_i$) is the vector of degrees (resp. total powers and final powers) of the vertices in component i . Using Lemma 5.2 we bound $|\mathcal{S}(k_i, \ell_i, 1, \bar{d}|_i, \bar{D}|_i, \bar{\delta}|_i)| \leq \frac{2^{qk_i} k_i^{\eta k_i}}{\prod_{v \in C_i} d_v!}$. Therefore, the total number of bits used to encode the hyperedges is at most

$$\left\lceil \log_2 \left(\left(\prod_v \frac{1}{d_v!} \right) 2^{qk} \left(\prod_{i=1}^c k_i^{\eta k_i} \right) \right) \right\rceil \quad (5.24)$$

bits.

Combining, (5.21), (5.22), (5.23), (5.24) we obtain that the total number of bits used to encode an arbitrary hypergraph in $\mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})$ is upper bounded by the maximum over feasible C_1, \dots, C_c and k_1, \dots, k_c of

$$\ell + c + \sum_{i=1}^c \left\lceil \log_2 \binom{\ell}{\ell_i - 1} \right\rceil + k + c + \left\lceil \log_2 \left(\frac{k!}{k_1! \dots k_c!} \right) \right\rceil + \left\lceil \log_2 \left(\left(\prod_v \frac{1}{d_v!} \right) 2^{qk} \left(\prod_{i=1}^c k_i^{\eta k_i} \right) \right) \right\rceil,$$

which we'll denote by b . We can safely add trailing zeros so that each hypergraph is encoded using exactly b bits. We conclude that the number of hypergraphs is at most 2^b , which is at most the right-hand side of the Lemma statement, as desired. \blacksquare

Lemma 5.4 *For any $k, \Gamma, \ell, c, q \geq \eta \geq 1, \bar{D}, \bar{d} \geq \bar{2}$ and $\bar{\delta}$ we have*

$$\begin{aligned} & \left(\prod_v \frac{D_v!}{\delta_v!} \right) |\mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})| \\ & \leq e^{O(qk)} k^{qk - \sum_{v \in [\ell]} (\delta_v - 1)} \Gamma^{qk - \ell - \sum_{v \in [\ell]} \delta_v} \max_{\substack{C_1, \dots, C_c \\ k_1, \dots, k_c \\ \text{feasible}}} \prod_{i=1}^c \left(\frac{k_i}{k} \right)^{(q-1)k_i - \sum_{v \in C_i} (\delta_v - 1) - (|C_i| - 1)}. \end{aligned} \quad (5.25)$$

Proof. Applying Lemma 5.3 we get

$$\begin{aligned} & \left(\prod_v \frac{D_v!}{\delta_v!} \right) |\mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})| \\ & \leq \max_{\substack{C_1, \dots, C_c \\ k_1, \dots, k_c}} \left[\left(2^{\ell + k + 5c + qk} k! \right) \cdot \left(\prod_{v=1}^{\ell} \frac{D_v!}{d_v! \delta_v!} \right) \cdot \left(\prod_{i=1}^c \binom{\ell}{|C_i| - 1} \right) \cdot \left(\prod_{i=1}^c \frac{k_i^{\eta k_i}}{k_i!} \right) \right]. \end{aligned} \quad (5.26)$$

We will now bound each factor of (5.26) in turn, making frequent use of the formula $n! = (n/e)^n n^{-O(1)}$ and the inequality $\binom{n}{m} \leq n^m/m!$.

First we bound

$$2^{\ell+k+5c+qk} k! \leq e^{O(qk)} k^k. \quad (5.27)$$

Secondly we bound

$$\begin{aligned} \prod_{v=1}^{\ell} \frac{D_v!}{d_v! \delta_v!} &= \prod_{i=1}^c \prod_{v \in C_i} \frac{(D_v - d_v)!}{\delta_v!} \cdot \binom{D_v}{d_v} \\ &\leq \prod_{i=1}^c \prod_{v \in C_i} D_v^{\max(0, D_v - d_v - \delta_v)} 2^{D_v} \\ &\leq \prod_{i=1}^c \prod_{v \in C_i} D_v^{D_v - d_v - \delta_v + 1} 2^{D_v} \\ &\leq \prod_{i=1}^c (\Gamma k_i)^{qk_i - \eta k_i - \sum_{v \in C_i} (\delta_v - 1)} 2^{qk_i} \\ &= \Gamma^{(q-\eta)k - \sum_{v \in [\ell]} (\delta_v - 1)} 2^{qk} \prod_{i=1}^c k_i^{(q-\eta)k_i - \sum_{v \in C_i} (\delta_v - 1)} \end{aligned}$$

using the facts $\sum_{v \in C_i} D_v = qk_i$, $\sum_{v \in C_i} d_v = \eta k_i$ and $D_v \leq \Gamma k_i \leq qk_i$. We finally observe that $(q-\eta)k - \sum_{v \in [\ell]} (\delta_v - 1) \leq qk - 2\ell - \sum_{v \in [\ell]} \delta_v + \ell$, yielding

$$\prod_{v=1}^{\ell} \frac{D_v!}{d_v! \delta_v!} \leq \Gamma^{qk - \ell - \sum_{v \in [\ell]} \delta_v} 2^{qk} \prod_{i=1}^c k_i^{(q-\eta)k_i - \sum_{v \in C_i} (\delta_v - 1)}. \quad (5.28)$$

For the third factor we consider two cases. If $\eta \geq 2$ we have $\ell_i \geq 2$ and $c \leq \ell/2$ and we bound

$$\begin{aligned} \prod_{i=1}^c \binom{\ell}{\ell_i - 1} &\leq \prod_{i=1}^c \frac{\ell^{\ell_i - 1}}{(\ell_i - 1)!} \leq \prod_{i=1}^c \frac{e^{O(\ell_i)} \ell^{\ell_i - 1}}{(\ell_i - 1)^{\ell_i - 1}} \\ &= \prod_{i=1}^c \frac{e^{O(\ell_i)} (\ell - c)^{\ell_i - 1}}{(\ell_i - 1)^{\ell_i - 1}} = e^{O(\eta k)} \prod_{i=1}^c \left(\frac{\ell_i - 1}{\ell - c} \right)^{-(\ell_i - 1)} \\ &\leq e^{O(\eta k)} \prod_{i=1}^c \left(\frac{k_i}{k} \right)^{-(\ell_i - 1)} \end{aligned} \quad (5.29)$$

where the last inequality is Gibbs'. It turns out that (5.29) holds when $\eta = 1$ as well because every $\ell_i = 1$ and hence both $\prod_{i=1}^c \binom{\ell}{\ell_i - 1}$ and $\prod_{i=1}^c \left(\frac{k_i}{k} \right)^{-(\ell_i - 1)}$ are equal to 1.

Fourth we write

$$\prod_{i=1}^c \frac{k_i^{\eta k_i}}{k_i!} = e^{O(\eta k)} \prod_{i=1}^c k_i^{(\eta - 1)k_i}. \quad (5.30)$$

Combining (5.27), (5.28), (5.29), (5.30) and (5.26) we get

$$\begin{aligned}
& \left(\prod_v \frac{D_v!}{\delta_v!} \right) |\mathcal{S}(\ell, c, \bar{d}, \bar{D}, \bar{\delta})| \\
& \leq \max_{\substack{C_1, \dots, C_c \\ k_1, \dots, k_c}} \left[\left(e^{O(qk)} k^k \right) \cdot \left(\Gamma^{qk-\ell-\sum_{v \in [\ell]} \delta_v} 2^{qk} \prod_{i=1}^c k_i^{(q-\eta)k_i - \sum_{v \in C_i} (\delta_v - 1)} \right) \right. \\
& \quad \cdot \left. \left(e^{O(\eta k)} \prod_{i=1}^c \left(\frac{k_i}{k} \right)^{-(\ell_i - 1)} \right) \cdot \left(e^{O(\eta k)} \prod_{i=1}^c k_i^{(\eta-1)k_i} \right) \right] \\
& = e^{O(qk)} k^k \Gamma^{qk-\ell-\sum_{v \in [\ell]} \delta_v} \max_{\substack{C_1, \dots, C_c \\ k_1, \dots, k_c}} \prod_{i=1}^c k_i^{(q-1)k_i - \sum_{v \in C_i} (\delta_v - 1)} \left(\frac{k_i}{k} \right)^{-(\ell_i - 1)} \\
& = e^{O(qk)} k^{qk - \sum_{v \in [\ell]} (\delta_v - 1)} \Gamma^{qk-\ell-\sum_{v \in [\ell]} \delta_v} \max_{\substack{C_1, \dots, C_c \\ k_1, \dots, k_c}} \prod_{i=1}^c \left(\frac{k_i}{k} \right)^{(q-1)k_i - \sum_{v \in C_i} (\delta_v - 1) - (\ell_i - 1)}.
\end{aligned}$$

■

Our final counting lemma bounds the optimization problem of Lemma 5.4.

Lemma 5.5 *For any $k, \ell, c, q \geq \eta \geq 1$, $\bar{d} \geq \bar{2}$, \bar{D} , and $\bar{\delta}$ we have*

$$\max_{\substack{C_1, \dots, C_c \\ k_1, \dots, k_c \\ \text{feasible}}} \prod_{i=1}^c \left(\frac{k_i}{k} \right)^{(q-1)k_i - \sum_{v \in C_i} (\delta_v - 1) - (|C_i| - 1)} \leq k^{-(c-1)(q-1)} e^{O(qk)}.$$

Proof. We are looking to upper-bound

$$\max_{\substack{\alpha_1, \dots, \alpha_c \\ z_1, \dots, z_c \\ \text{feasible}}} \prod_{i=1}^c \alpha_i^{z_i} \equiv \mathcal{M} \tag{5.31}$$

where $z_i = (q-1)k_i - (|C_i| - 1) - \sum_{v \in C_i} (\delta_v - 1) = 1 + (q-1)k_i - \sum_{v \in C_i} \delta_v$ and $\alpha_i = k_i/k$.

We upper-bound \mathcal{M} by the relaxation

$$\max_{\substack{\alpha_1, \dots, \alpha_c \\ z_1, \dots, z_c}} \prod_{i=1}^c \alpha_i^{z_i} \text{ such that} \tag{5.32}$$

$$\sum_i \alpha_i = 1 \tag{5.33}$$

$$\alpha_i \geq 0 \tag{5.34}$$

$$\sum_i z_i = Z \tag{5.35}$$

$$z_i \geq q-1 \tag{5.36}$$

where $Z = \sum_i z_i = c + (q-1)k - \sum_v \delta_v$. To show this is a relaxation we need to prove that any z_i feasible in (5.31) satisfies (5.36), which follows from Lemma 2.5 which states that $(q-1)k_i - \sum_{v \in C_i} \delta_v \geq q-2$. Another implication of that lemma is that $Z \geq (q-1)c$.

When $q = 1$ we can trivially prove the Lemma by upper-bounding (5.32) by 1, so we hereafter assume $q \geq 2$ and hence every z_i is strictly positive. For any fixed $\{z_i\}_i$, Gibbs' inequality implies that the maximum of (5.32) occurs when $\alpha_i = z_i/Z$. Therefore we have reduced our problem to

$$\begin{aligned} \max_{z_1, \dots, z_c} \prod_{i=1}^c \left(\frac{z_i}{Z}\right)^{z_i} \quad \text{such that} \\ \sum_i z_i = Z \\ z_i \geq q - 1. \end{aligned} \tag{5.37}$$

Clearly the optimum is when $z_i = q - 1$ for all $i \neq 1$ and $z_1 = Z - (c - 1)(q - 1)$. Therefore the maximum of (5.37) is

$$\begin{aligned} \left(\frac{Z - (c - 1)(q - 1)}{Z}\right)^{Z - (c - 1)(q - 1)} \left(\frac{q - 1}{Z}\right)^{(c - 1)(q - 1)} &\leq 1 \cdot \left(\frac{1}{c}\right)^{(c - 1)(q - 1)} \\ &= \left(\frac{1}{k}\right)^{(c - 1)(q - 1)} \left(\frac{k}{c}\right)^{(c - 1)(q - 1)} \\ &\leq k^{-(c - 1)(q - 1)} e^{O(qk)} \end{aligned} \tag{5.38}$$

using the fact that $Z = \sum_i z_i \geq (q - 1)c$ in the first inequality. ■

We are finally ready to prove our Main Counting Lemma.

Proof. of Lemma 5.1. Lemmas 5.4 and 5.5 give us

$$\begin{aligned} &\left(\prod_v \frac{D_v!}{\delta_v!}\right) |\mathcal{S}(k, \ell, c, \bar{d}, \bar{D}, \bar{\delta})| \\ &\leq e^{O(qk)} k^{qk - \sum_{v \in [\ell]} (\delta_v - 1)} \Gamma^{qk - \ell - \sum_{v \in [\ell]} \delta_v} k^{-c(q - 1) + (q - 1)} \\ &= e^{O(qk)} \Gamma^{qk - \ell - \sum_{v \in [\ell]} \delta_v} k^{qk - c(q - 1) - \sum_{v \in [\ell]} (\delta_v - 1)} \\ &\leq R_0^{qk} \Gamma^{qk - \ell - \sum_{v \in [\ell]} \delta_v} k^{qk - c(q - 1) - \sum_{v \in [\ell]} (\delta_v - 1)} \end{aligned}$$

for some absolute constant $R_0 > 1$ (we used the fact that $k^{q-1} = e^{O(qk)}$). ■

6 Permanents of Random Matrices

Proof. of Theorem 1.5 Notice first that $\mu_t \leq (n - t)! \leq n^{n-t}$ and the power of the polynomial $q = n$. Since the permanent is a multilinear polynomial and $\mathbb{E}[Y_{ij}] = 0$ we can directly apply the Lemma 2.1 for $k \leq n$. Note also that n in this Theorem is the dimension of the matrix and not the number of random variables as in Lemma 2.1 (which is n^2 in this setting). We obtain

$$\begin{aligned} \left|\mathbb{E}\left[P(A)^k\right]\right| &\leq \max_{\ell, \bar{\nu}} \left\{ R^{nk} k^{nk - \ell} \prod_{t=0}^n n^{(n-t)\nu_t} \right\} \\ &= \max_{\ell} \left\{ R^{nk} k^{nk - \ell} n^{\ell} \right\} = R^{nk} k^{nk} \max_{\ell} \left\{ \left(\frac{n}{k}\right)^{\ell} \right\} \leq R^{nk} k^{nk/2} n^{nk/2}. \end{aligned}$$

We fix the deviation $\lambda = t\sqrt{n!} > 0$ and choose k^* to be the even number in the interval $(K - 2, K]$ for $K = \frac{(\lambda/e)^{2/n}}{R^2 n}$. Using the Markov's inequality we derive

$$\begin{aligned} \Pr[|P(A)| \geq \lambda] &\leq \frac{\mathbb{E}[|P(A)|^{k^*}]}{\lambda^{k^*}} \\ &\leq e^{k^* \ln \frac{R^n (k^*)^{n/2} n^{n/2}}{\lambda}} \leq e^{-k^*} \leq e^{-K+2} \\ &\leq e^2 \cdot e^{-\frac{(\lambda/e)^{2/n}}{R^2 n}} \leq e^2 \cdot e^{-ct^{2/n}}, \end{aligned}$$

for some absolute constant $c > 0$. Note that condition $k^* \leq n$ is implied by the condition $K \leq n$ which in turn is equivalent to the condition $\lambda \leq eR^n n^n$. If $\lambda > eR^n n^n$ then we choose $k^* = n$ and estimate

$$\begin{aligned} \Pr[|P(A)| \geq \lambda] &< \Pr[|P(A)| \geq eR^n n^n] \\ &\leq \frac{\mathbb{E}[|P(A)|^{k^*}]}{(eR^n n^n)^{k^*}} \leq e^{-n} \end{aligned}$$

■

Proof. of Theorem 1.6 We have an n by n matrix A with entries that are independent except that the matrix is symmetric. The permanent $P(A)$ is a degree n polynomial of independent random variables with maximal variable degree $\Gamma = 2$. (Note that the number of variables is $\binom{n}{2} + n$, not n .) The permanent is a sum of products over permutations. We also treat each such permutation π as a set of pairs (i, j) for row index i and column index j . We write $h(\pi)$ for the hyperedge h corresponding to π . More generally for any set S of matrix entries we write $h(S)$ for the corresponding hyperedge. Note that because each variable appears in up to two positions in the matrix the mapping h is not a bijection. Clearly

$$\begin{aligned} P(A) &= \sum_{\pi} \prod_{i=1}^n A_{i, \pi(i)} = \sum_h \sum_{\pi: h(\pi)=h} \prod_{i=1}^n A_{i, \pi(i)} \\ &= \sum_h \underbrace{\left(\sum_{\pi: h(\pi)=h} 1 \right)}_{\equiv w_h} \left(\prod_{v \in \mathcal{V}(h)} Y_v^{\tau_{hv}} \right) \end{aligned}$$

As in the proof of Lemma 3.3 we write $P(A)$ as the sum of polynomials $g^{(1)}, \dots, g^{(m)}$ with weights $w^{(1)}, \dots, w^{(m)}$ and total powers q_1, \dots, q_m (in this case $\mathbb{E}[P(A)] = 0$).

Fix some $1 \leq i \leq m$ and hyperedge h' with $q(h') = q_i$. The next step is to bound coefficients of polynomials $g^{(i)}$,

$$\begin{aligned} w_{h'}^{(i)} &= \sum_{h \supseteq h'} w_h \left(\prod_{v \in \mathcal{V}(h) \setminus \mathcal{V}(h')} \mathbb{E}[Y_v^{\tau_{hv}}] \right) \\ &= \sum_{\pi: h(\pi) \supseteq h'} \left(\prod_{v \in \mathcal{V}(h(\pi)) \setminus \mathcal{V}(h')} \mathbb{E}[Y_v^{\tau_{hv}}] \right) \\ &\leq \sum_{S: h(S)=h'} \sum_{\pi: \pi \supseteq S} \left(\prod_{v \in \mathcal{V}(h(\pi)) \setminus \mathcal{V}(h')} \mathbb{E}[Y_v^{\tau_{hv}}] \right) \end{aligned}$$

where S is a set matrix entries. We can bound the number of such S by $2^{O(q_i)}$ since for each vertex (variable) $v \in h'$ there are only two entries in the matrix A that can be mapped to v by the mapping $h(S)$. Fix an S such that $h(S) = h'$. Note that whenever $\tau_{hv} = 1$ we have $\mathbb{E}[Y_v^{\tau_{hv}}] = 0$, so we can restrict the sum to be over permutations π with $\tau_{hv} = 2$ for all $v \in \mathcal{V}(h(\pi)) \setminus \mathcal{V}(h')$. For every fixed S , the number of such π is at most $n^{(n-q_i)/2}$ since we need to choose the remaining $n - q_i$ entries and each choice fixes two positions in π . By moment boundedness we have $\left(\prod_{v \in \mathcal{V}(\pi) \setminus \mathcal{V}(h')} \mathbb{E}[Y_v^{\tau_{hv}}]\right) \leq 2^{(n-q_i)/2}$. We conclude that $w_{h'}^{(i)} \leq 2^{O(n)} n^{(n-q_i)/2}$.

Fix some $g^{(i)}$ with total power $q_i \equiv q \leq n$. We start from (2.11):

$$\begin{aligned}
\left| \mathbb{E} \left[g^{(i)}(Y)^k \right] \right| &\leq \sum_{\ell=\eta}^{k\eta/2} \frac{1}{\ell!} \sum_{G' \in \mathcal{S}_2(k, \ell)} \underbrace{\sum_{\pi \in M([\ell])}}_{\leq (n^2)^\ell} \underbrace{\left(\prod_{h \in \mathcal{H}(G')} w_{\pi(h)} \right)}_{\leq (2^{O(n)} n^{(n-q)/2})^k} \underbrace{\left(\prod_{u \in \mathcal{V}(G')} \Lambda_{\pi(u)}(G') \right)}_{\leq 2^{qk} \prod_u D_u!} \\
&\leq \sum_{\ell=\eta}^{k\eta/2} \frac{1}{\ell!} \sum_{G' \in \mathcal{S}_2(k, \ell)} (n^2)^\ell (2^{O(n)} n^{(n-q)/2})^k \left(2^{qk} \prod_{u \in \mathcal{V}(G')} D_u! \right) \\
&\leq \sum_{\ell=\eta}^{k\eta/2} 2^{O(n)} R_6^{qk} \frac{1}{\ell!} (n^2)^\ell k^{qk} n^{k(n-q)/2} \\
&\leq R_7^{nk} \left(\frac{n^{2\ell}}{\ell^\ell} \right) k^{qk} n^{k(n-q)/2} \tag{6.39}
\end{aligned}$$

where the third inequality bounded $2^{qk} |\mathcal{S}_2(\ell)| \prod_u D_u!$ by $R_6^{qk} k^{qk}$ using Lemma 5.1. Recall that $\max_{x>0} (nk/x)^x = e^{nk/e}$. We continue using the facts that we choose $k \leq n$ and $\ell \leq \eta k/2 \leq qk/2$,

$$\begin{aligned}
\left| \mathbb{E} \left[g^{(i)}(Y)^k \right] \right| &\leq R_7^{nk} \left(\frac{nk}{\ell} \right)^\ell \left(\frac{n}{k} \right)^\ell k^{qk} n^{k(n-q)/2} \\
&\leq R_8^{nk} \left(\frac{n}{k} \right)^\ell k^{qk} n^{k(n-q)/2} \\
&\leq R_8^{nk} n^{qk/2} k^{qk/2} n^{k(n-q)/2} \\
&\leq R_8^{nk} n^{nk/2} k^{nk/2}.
\end{aligned}$$

Applying Corollary 3.2 yields

$$\begin{aligned}
\mathbb{E} \left[|P(A)|^k \right] &\leq \mathbb{E} \left[\left(\sum_{i=1}^m |g^{(i)}(Y)| \right)^k \right] \\
&\leq m^k R_8^{nk} n^{nk/2} k^{nk/2} \\
&\leq R_9^{nk} n^{nk/2} k^{nk/2}
\end{aligned}$$

since $m \leq 2n^2 \leq 10^n$ and where R_9 is an absolute constant.

This gives us a bound on $|\mathbb{E}[P(A)^k]|$ identical to that in the proof of Theorem 1.5, so we finish the proof identically to the proof of Theorem 1.5. \blacksquare

7 Examples of Moment Bounded Random Variables

In this section we show that three classes of random variables are moment bounded and give examples from each class. The classes are bounded random variables, log-concave continuous

random variables, and log-concave discrete random variables.

7.1 Bounded random variables

Lemma 7.1 *Any random variable Z with $|Z| \leq L$ is moment bounded with parameter L .*

Proof. For any $i \geq 1$ we clearly have $|Z|^i \leq L|Z|^{i-1}$ hence $\mathbb{E}[|Z|^i] \leq L\mathbb{E}[|Z|^{i-1}] \leq iL\mathbb{E}[|Z|^{i-1}]$. ■

In particular Lemma 7.1 implies that $0/1$ and $-1/1$ random variables are moment bounded with parameter 1.

7.2 Log-concave continuous random variables

We say that non-negative function f is *log-concave* if $f(\lambda x + (1 - \lambda)y) \geq f(x)^\lambda f(y)^{1-\lambda}$ for any $0 \leq \lambda \leq 1$ and $x, y \in \mathbb{R}$ (see [15] Section 3.5). Equivalently f is log concave if $\ln f(x)$ is concave on the set $\{x : f(x) > 0\}$ where $\ln f(x)$ is defined and this set is a convex set (i.e. an interval). A continuous random variable (or a continuous distribution) with density f is *log-concave* if f is a log-concave function. See [5, 6, 8, 15] for introductions to log-concavity.

Lemma 7.2 *Any non-negative log-concave random variable X with density f is moment bounded with parameter $L = \mathbb{E}[X]$.*

Proof. Let $\ell = \inf\{x \geq 0 : f(x) > 0\}$ and $u = \sup\{x \geq 0 : f(x) > 0\}$. By log-concavity we have that $f(x) > 0$ for all $\ell < x < u$. Let $F(x) = \Pr[X \leq x]$ and $\bar{F}(x) = \Pr[X \geq x]$. Note that $\bar{F}(x) = 0$ for all $x \geq u$. For any $i \geq 1$ we write

$$\begin{aligned} \mathbb{E}[X^i] &= \int_{x=0}^{\infty} x^i dF(x) \\ &= - \int_{x=0}^{\infty} x^i d\bar{F}(x) \\ &= -x^i \bar{F}(x) \Big|_{x=0}^{\infty} + \int_{x=0}^{\infty} \bar{F}(x) d(x^i) \\ &= 0 + \int_0^{\infty} \bar{F}(x) i x^{i-1} dx \\ &= \int_0^{\ell} i x^{i-1} dx + \int_{\ell}^u \frac{\bar{F}(x)}{f(x)} i x^{i-1} f(x) dx \end{aligned} \tag{7.40}$$

where the third equality is integration by parts. It is known (see for example implication B of Proposition 1 in [5] or Theorem 2 in [8]) that log-concavity of density f implies log-concavity of $\bar{F}(x)$. It follows that $d(\ln \bar{F}(x))/dx = -f(x)/\bar{F}(x)$ is a non-increasing function of x on (ℓ, u) and hence $\bar{F}(x)/f(x)$ is also non-increasing. It follows that $\frac{\bar{F}(x)}{f(x)} \cdot i x^{i-1}$ is a product of a non-increasing

function and a non-decreasing function. We apply Chebyshev's integral inequality, yielding,

$$\begin{aligned}
\mathbb{E}[X^i] &\leq \int_0^\ell ix^{i-1}dx + \left[\int_\ell^u \frac{\bar{F}(x)}{f(x)} f(x) dx \right] \left[\int_\ell^u ix^{i-1} f(x) dx \right] \\
&= \int_0^\ell ix^{i-1}dx + \left[\int_\ell^u \bar{F}(x) dx \right] \cdot i\mathbb{E}[X^{i-1}] \\
&= \ell^i + \left[\int_\ell^u \bar{F}(x) dx \right] \cdot i\mathbb{E}[X^{i-1}] \\
&\leq i\ell\mathbb{E}[X^{i-1}] + \left[\int_\ell^u \bar{F}(x) dx \right] \cdot i\mathbb{E}[X^{i-1}] \\
&= \left[\int_0^{+\infty} \bar{F}(x) dx \right] \cdot i\mathbb{E}[X^{i-1}] = \mathbb{E}[X] \cdot i\mathbb{E}[X^{i-1}]
\end{aligned}$$

where we used the fact $\mathbb{E}[|X|] = \int_0^{+\infty} \bar{F}(x) dx$. ■

Lemma 7.3 *Any log-concave random variable X with density f is moment bounded with parameter $L = \frac{1}{\ln 2} \mathbb{E}[|X|] \approx 1.44 \mathbb{E}[|X|]$.*

Proof. If X is non-negative or non-positive with probability 1 the Lemma follows from Lemma 7.2, so suppose not. Write

$$\begin{aligned}
\mathbb{E}[|X|^k] &= \Pr[X \geq 0] \mathbb{E}[X^k | X \geq 0] + \Pr[X < 0] \mathbb{E}[(-X)^k | X < 0] \\
&= \Pr[X \geq 0] \mathbb{E}[X_+^k] + \Pr[X < 0] \mathbb{E}[X_-^k]
\end{aligned}$$

where X_+ (resp. X_-) is a non-negative random variable with density at x proportional to $f(x)$ (resp. $f(-x)$) for $x \geq 0$ and zero for $x < 0$. Clearly X_+ and X_- are log-concave. Lemma 7.2 yields

$$\begin{aligned}
\mathbb{E}[|X|^k] &\leq \Pr[X \geq 0] k\mathbb{E}[|X_+|] \mathbb{E}[|X_+|^{k-1}] + \Pr[X < 0] k\mathbb{E}[|X_-|] \mathbb{E}[|X_-|^{k-1}] \\
&\leq k \max\{\mathbb{E}[|X_+|], \mathbb{E}[|X_-|]\} \left(\Pr[X \geq 0] \mathbb{E}[|X|^{k-1} | X \geq 0] + \Pr[X < 0] \mathbb{E}[|X|^{k-1} | X < 0] \right) \\
&= k \max\{\mathbb{E}[|X| | X \geq 0], \mathbb{E}[|X| | X < 0]\} \mathbb{E}[|X|^{k-1}] \\
&\leq k \frac{1}{\ln 2} \mathbb{E}[|X|] \mathbb{E}[|X|^{k-1}]
\end{aligned}$$

where we used Lemma 7.4 in the last inequality to bound $\max\{\mathbb{E}[|X| | X \leq 0], \mathbb{E}[|X| | X \geq 0]\} \leq \frac{1}{\ln 2} \mathbb{E}[|X|]$. ■

The survey [8] lists many distributions with log-concave densities: normal, exponential, logistic, extreme value, chi-square, chi, Laplace, Weibull, Gamma, and Beta, where the last three are log-concave only for some parameter values. Lemma 7.3 implies that random variables with any of these distributions are moment bounded.

Any random variable trivially satisfies $\mathbb{E}[|X|^1] = 1\mathbb{E}[|X|] \mathbb{E}[|X|^{1-1}]$ so for every random variable (that is non-zero with positive probability) Lemma 7.2 gives the smallest possible moment boundedness parameter L and Lemma 7.3 gives L that is within a factor of $1/\ln 2$ of the best possible. An exponentially distributed random variable is tight for Lemma 7.2 in an even stronger sense: $\mathbb{E}[|X|^k] = k\mathbb{E}[|X|] \mathbb{E}[|X|^{k-1}]$ for all integers $k \geq 1$.

The following example shows that Lemma 7.3 is tight. Let X have density

$$f(x) = \begin{cases} e^{(x-x_0)} & \text{if } x \leq x_0 \\ 0 & \text{if } x > x_0 \end{cases}$$

where $x_0 = \ln 2$. This density is clearly log-concave. Using integration by parts we derive

$$\begin{aligned} \mathbb{E}[|X|] &= \int_0^{+\infty} x e^{-x-x_0} dx + \int_0^{\ln 2} x e^{x-x_0} dx \\ &= -\frac{x e^{-x}}{2} \Big|_0^{+\infty} + \int_0^{+\infty} \frac{e^{-x}}{2} dx + \frac{x e^x}{2} \Big|_0^{\ln 2} - \int_0^{\ln 2} \frac{e^x}{2} dx \\ &= 0 + 1/2 + \ln 2 - 1/2 = \ln 2. \end{aligned}$$

The k -th moment for large k is dominated by the exponential left tail: $\mathbb{E}[|X|^k] = e^{-x_0} k! + O(1)$. Therefore $\lim_{k \rightarrow \infty} \frac{\mathbb{E}[|X|^k]}{\mathbb{E}[|X|^{k-1}]} = k = k \frac{1}{\ln 2} \mathbb{E}[|X|]$, hence X is moment-bounded for no $L < \frac{1}{\ln 2} \mathbb{E}[|X|]$.

7.2.1 Technical lemmas

This section is devoted to proving the following Lemma.

Lemma 7.4 *For any random variable X with log-concave density, $\Pr[X \geq 0] > 0$ and $\Pr[X \leq 0] > 0$ we have*

$$\max\{\mathbb{E}[|X| \mid X \leq 0], \mathbb{E}[|X| \mid X \geq 0]\} \leq \frac{1}{\ln 2} \mathbb{E}[|X|] \approx 1.44 \mathbb{E}[|X|].$$

The following Lemma about log-concave functions is intuitive but a bit technical to prove.

Lemma 7.5 *Suppose f is a log-concave function, $x_0 > 0$, $h(x) = \begin{cases} e^{x-x_0} & \text{if } x \leq x_0 \\ 0 & \text{if } x > x_0 \end{cases}$, $f(0) = h(0)$,*

and $\int_{-\infty}^0 f(x) dx = \int_{-\infty}^0 h(x) dx$. It follows that:

1. *there exists $x_1 < 0$ such that $(x - x_1)(f(x) - h(x)) \geq 0$ for any $x \leq 0$ and*
2. *$(x - x_0)(f(x) - h(x)) \geq 0$ for any $x \geq 0$.*

Proof. Let $S^+ = \{x \leq 0 : f(x) > h(x)\}$ and $S^- = \{x \leq 0 : f(x) < h(x)\}$. We have

$$0 = \int_{-\infty}^0 f(x) dx - \int_{-\infty}^0 h(x) dx = \int_{S^+} (f(x) - h(x)) dx - \int_{S^-} (h(x) - f(x)) dx$$

and an integral of a positive function is positive iff it is over a set of positive measure, hence either both S^+ and S^- have Lebesgue measure zero or neither do. Below we will use the following simple fact about concave functions. If $g(x)$ is concave, $g(z) = 0$, $g(z') > 0$ and $z' < z$ then $g(z'') > 0$ for all $z'' \in (z, z')$.

The log-concavity of f implies that $\ln f(x) - \ln h(x)$ is concave on $(-\infty, x_0]$. This and the fact that $f(0) = h(0)$ imply the following key properties: if $x \in S^+$ then $S^+ \supseteq [x, 0)$ and similarly if $y \in S^-$ then $S^- \supseteq (-\infty, y]$. Among other things these properties imply that if S^+ (resp. S^-) is non-empty it contains an interval and hence has positive measure. If both S^+ and S^- have measure zero then $f(x) = h(x)$ for all $x \leq 0$ and any $x_1 < 0$ will satisfy the first part of the lemma. If both S^+ and S^- have positive measure then $x_1 = \inf S^+$ will satisfy the first part of the lemma.

The first part of the lemma implies that $x_2 = x_1/2 < 0$ satisfies $f(x_2) \geq h(x_2)$. For any $0 < x < x_0$ the facts that $\ln f(x) - \ln h(x)$ is concave on $(-\infty, x_0]$, $f(x_2) \geq h(x_2)$ and $f(0) = h(0)$ imply that $f(x) \leq h(x)$. Clearly $f(x) \geq h(x) = 0$ for $x > x_0$, so the second part of the Lemma follows. ■

Now we prove Lemma 7.4.

Proof. Let random variable X be given with density f . We prove the upper-bound on $\mathbb{E}[|X| \mid X \leq 0]$ only; the upper-bound on $\mathbb{E}[|X| \mid X \geq 0]$ follows from this bound because $-X$ is log-concave. The Lemma is invariant with respect to scaling X so we can and do assume without loss of generality that $f(0) = \Pr[X \leq 0]$.

Let X' be a random variable with density $h(x) = \begin{cases} e^{x-x_0} & \text{if } x \leq x_0 \\ 0 & \text{if } x > x_0 \end{cases}$ where x_0 is the solution to $e^{-x_0} = \Pr[X \leq 0]$. One can readily verify that $\Pr[X' \leq 0] = e^{-x_0} = \Pr[X \leq 0]$. Therefore, $\Pr[X' \geq 0] = \Pr[X \geq 0]$, and $h(0) = f(0)$.

Using the first part of Lemma 7.5 we have

$$\begin{aligned} 0 &\leq \int_{-\infty}^0 (x - x_1)(f(x) - h(x))dx \\ &= \int_{-\infty}^0 x(f(x) - h(x))dx - x_1(\Pr[X \leq 0] - \Pr[X' \leq 0]) \\ &= \Pr[X \leq 0](-\mathbb{E}[|X| \mid X \leq 0] + \mathbb{E}[|X'| \mid X' \leq 0]) - 0 \end{aligned}$$

i.e. $\mathbb{E}[|X'| \mid X' \leq 0] \geq \mathbb{E}[|X| \mid X \leq 0]$.

By the second part of Lemma 7.5 we have

$$\begin{aligned} 0 &\leq \int_0^{+\infty} (x - x_0)(f(x) - h(x))dx \\ &= \int_0^{+\infty} x(f(x) - h(x))dx - x_0(\Pr[X \geq 0] - \Pr[X' \geq 0]) \\ &= \Pr[X \geq 0](\mathbb{E}[|X| \mid X \geq 0] - \mathbb{E}[|X'| \mid X' \geq 0]) - 0 \end{aligned}$$

i.e. $\mathbb{E}[|X'| \mid X' \geq 0] \leq \mathbb{E}[|X| \mid X \geq 0]$.

We conclude that

$$\begin{aligned} \frac{\mathbb{E}[|X|]}{\mathbb{E}[|X| \mid X \leq 0]} &= \frac{\Pr[X \leq 0] \mathbb{E}[|X| \mid X \leq 0] + \Pr[X \geq 0] \mathbb{E}[|X| \mid X \geq 0]}{\mathbb{E}[|X| \mid X \leq 0]} \\ &= \Pr[X \leq 0] + \Pr[X \geq 0] \cdot \frac{\mathbb{E}[|X| \mid X \geq 0]}{\mathbb{E}[|X| \mid X \leq 0]} \\ &\geq \Pr[X' \leq 0] + \Pr[X' \geq 0] \cdot \frac{\mathbb{E}[|X'| \mid X' \geq 0]}{\mathbb{E}[|X'| \mid X' \leq 0]} \\ &= \Pr[X' \leq 0] + \Pr[X' \geq 0] \frac{\Pr[X' \geq 0] \mathbb{E}[|X'| \mid X' \geq 0]}{\Pr[X' \leq 0] \mathbb{E}[|X'| \mid X' \leq 0]} \\ &= e^{-x_0} + e^{-x_0} \frac{x_0 + e^{-x_0} - 1}{e^{-x_0}} \\ &= x_0 + 2e^{-x_0} - 1 \equiv r(x_0) \end{aligned} \tag{7.41}$$

where we used integration by parts to compute

$$\begin{aligned}
\Pr[X' \leq 0] \mathbb{E}[|X'| \mid X' \leq 0] &= \int_{-\infty}^0 (-x) e^{x-x_0} dx \\
&= (-x) e^{x-x_0} \Big|_{-\infty}^0 - \int_{-\infty}^0 (-1) e^{x-x_0} dx = 0 + e^{-x_0} \\
\Pr[X' \geq 0] \mathbb{E}[|X'| \mid X' \geq 0] &= \int_0^{x_0} x e^{x-x_0} dx \\
&= x e^{x-x_0} \Big|_0^{x_0} - \int_0^{x_0} e^{x-x_0} dx = x_0 + e^{-x_0} - 1.
\end{aligned}$$

Finally we note that $r(x_0)$ is minimized on $0 \leq x_0 < \infty$ when $0 = \frac{dr}{dx_0} = 1 - 2e^{-x_0}$. We conclude that

$$\frac{\mathbb{E}[|X|]}{\mathbb{E}[|X| \mid X \leq 0]} \geq r(x_0) \geq r(\ln 2) = \ln 2$$

which implies the Lemma. ■

7.3 Log-concave discrete random variables

A distribution over the integers $\dots, p_{-2}, p_{-1}, p_0, p_1, p_2, \dots$ is said to be *log-concave* [5, 27] if $p_{i+1}^2 \geq p_i p_{i+2}$ for all i . An integer-valued random variable X is log-concave if its distribution $p_i = \Pr[X = i]$ is.

Lemma 7.6 *Any non-negative integer-valued log-concave random variable X is moment bounded with parameter $L = 1 + \mathbb{E}[|X|]$.*

Proof. The proof parallels the proof of Lemma 7.2. Let $r_i = \Pr[X \geq i] = \sum_{j=i}^{\infty} p_j$, $\ell = \min\{i : p_i > 0\}$ and $u = \max\{i : p_i > 0\}$. For any $k \geq 1$ we have

$$\begin{aligned}
\mathbb{E}[|X|^k] &= \sum_{x=0}^{\infty} p_x x^k \\
&= \sum_{x=1}^{\infty} (r_x - r_{x+1}) x^k \\
&= \sum_{x=1}^{\infty} r_x (x^k - (x-1)^k) \\
&\leq \sum_{x=1}^{\infty} r_x k x^{k-1} = \sum_{x=0}^{\infty} r_x k x^{k-1} \\
&= \sum_{x=0}^{\ell-1} k x^{k-1} + \sum_{x=\ell}^u \frac{r_x}{p_x} k x^{k-1} p_x \\
&\leq \sum_{x=0}^{\ell-1} k x^{k-1} + \left(\sum_{x=\ell}^u \frac{r_x}{p_x} p_x \right) \left(\sum_{x=\ell}^u k x^{k-1} p_x \right) \\
&\leq \max\{0, \ell-1\} \mathbb{E}[k|X|^{k-1}] + (\mathbb{E}[|X|] + 1 - \ell) \mathbb{E}[k|X|^{k-1}] \\
&= (1 + \mathbb{E}[|X|]) \mathbb{E}[k|X|^{k-1}]
\end{aligned}$$

where the second inequality uses the fact that $\frac{r_x}{p_x}$ is a non-increasing sequence (Proposition 10 in [5]) and Chebyshev's sum inequality. \blacksquare

Lemma 7.7 *Any log-concave integer-valued random variable X is moment bounded with parameter $L = \max(\mathbb{E}[|X| \mid X \geq 0], \mathbb{E}[|X| \mid X < 0])$.*

We omit the proof of Lemma 7.7, which is almost identical to the proof of Lemma 7.3.

Examples of log-concave integer-valued distributions include Poisson, binomial, negative binomial and hypergeometric [5, 27]. Random variables with these distributions are moment bounded by Lemma 7.6.

The parameter $1 + \mathbb{E}[|X|]$ in Lemma 7.6 cannot be improved to match the $\mathbb{E}[|X|]$ in Lemma 7.2. Indeed a Poisson distributed random variable with mean μ has $\mathbb{E}[X^2] = \mu^2 + \mu$, which exceeds the desired bound of $2\mathbb{E}[X]\mathbb{E}[X^{2-1}] = 2\mu^2$ when $\mu < 1$.

8 Examples Showing Tightness of the Bounds

This section deals exclusively with multilinear polynomials with non-negative coefficients over independent 0/1 random variables. We use notation specialized to this case: for a polynomial $f(x)$ and 0/1 random variables Y_1, \dots, Y_n we have

$$\mu_r(f, Y) = \max_{A \subseteq \mathcal{V}: |A|=r} \left\{ \sum_{h \in \mathcal{H} | A \subseteq h} w_h \prod_{i \in h \setminus A} \mathbb{E}[Y_i] \right\}.$$

We continue to omit the Y from $\mu_r(f, Y)$ when it is clear from context.

Lemma 8.1 *We are given a power q_1 polynomial $f_1(x)$ with corresponding hypergraph H_1 , weights w_1 , vertices $\mathcal{V}(H_1) = [n]$ and a power q_2 polynomial $f_2(x)$ with corresponding hypergraph H_2 , weights w_2 , vertices $\mathcal{V}(H_2) = \{n+1, \dots, m\}$. We are also given m independent 0/1 random variables X_1, \dots, X_m . Then the product polynomial $fg = (H, w)$ defined by $(fg)(x_1, \dots, x_m) = f(x_1, \dots, x_n)g(x_{n+1}, \dots, x_m)$ satisfies*

$$\mu_i(fg, X) = \max_{0 \leq i_1 \leq q_1, 0 \leq i_2 \leq q_2} \mu_{i_1}(f, X_1, \dots, X_n) \mu_{i-i_1}(g, X_{n+1}, \dots, X_m).$$

Proof. The Lemma follows easily from the definition of μ_i and the fact that restriction to hyperedges containing a fixed set of vertices preserves the product structure. Indeed let $\mathcal{H}_1 = \mathcal{H}(H_1)$, $\mathcal{H}_2 = \mathcal{H}(H_2)$ and $\mathcal{H} = \mathcal{H}(H)$. Then

$$\begin{aligned} \mu_i(fg) &= \max_{A \subseteq \mathcal{V}: |A|=r} \left\{ \sum_{h \in \mathcal{H} | A \subseteq h} w_h \prod_{v \in h \setminus A} \mathbb{E}[X_v] \right\} \\ &= \max_{\substack{0 \leq i_1 \leq q_1, 0 \leq i_2 \leq q_2: \\ i_1 + i_2 = i}} \max_{\substack{A_1 \subseteq \mathcal{V}_1: \\ |A_1|=i_1}} \max_{\substack{A_2 \subseteq \mathcal{V}_2: \\ |A_2|=i_2}} \left\{ \sum_{\substack{h_1 \in \mathcal{H}_1: \\ A_1 \subseteq h_1}} \sum_{\substack{h_2 \in \mathcal{H}_2: \\ A_2 \subseteq h_2}} w_{h_1} w_{h_2} \prod_{v \in h_1 \setminus A_1} \mathbb{E}[X_v] \prod_{v \in h_2 \setminus A_2} \mathbb{E}[X_v] \right\} \\ &= \max_{\substack{0 \leq i_1 \leq q_1, 0 \leq i_2 \leq q_2: \\ i_1 + i_2 = i}} \left(\max_{\substack{A_1 \subseteq \mathcal{V}_1: \\ |A_1|=i_1}} \sum_{\substack{h_1 \in \mathcal{H}_1: \\ A_1 \subseteq h_1}} \prod_{v \in h_1 \setminus A_1} \mathbb{E}[X_v] \right) \left(\max_{\substack{A_2 \subseteq \mathcal{V}_2: \\ |A_2|=i_2}} \sum_{\substack{h_2 \in \mathcal{H}_2: \\ A_2 \subseteq h_2}} \prod_{v \in h_2 \setminus A_2} \mathbb{E}[X_v] \right) \\ &= \max_{\substack{0 \leq i_1 \leq q_1, 0 \leq i_2 \leq q_2: \\ i_1 + i_2 = i}} \mu_{i_1}(f) \mu_{i_2}(g). \end{aligned}$$

■

Our next lemma studies a particular sort of complete multilinear q -uniform polynomials that we will use frequently.

Lemma 8.2 *Given $Z = \left(\sum_{i=1}^n X_i\right)_q = \sum_{h \subseteq [n]: |h|=q} \prod_{v \in h} X_v$ where the X_v are independent 0/1 random variables with $\mathbb{E}[X_i] = p \leq 0.5$ we have*

- $\mu_i(Z) = \binom{n-i}{q-i} p^{q-i} \leq (np)^{q-i}$ and
- $\Pr \left[Z = \binom{c}{q} \right] = \binom{n}{c} p^c (1-p)^{n-c} \geq e^{-2np} \left(\frac{np}{c}\right)^c$ for any integer $0 \leq c \leq n$.

Proof. The first is immediate from definitions. The second follows because

$$\Pr \left[Z = \binom{c}{q} \right] = \Pr \left[\sum_i X_i = c \right] = \binom{n}{c} p^c (1-p)^{n-c} \geq (n/c)^c p^c e^{(n-c) \ln(1-p)} \geq (np/c)^c e^{-2np}$$

where we used $\ln(1-p) \geq -2p$ for $0 \leq p \leq 1/2$ in the last inequality. ■

Lemma 8.3 *For any $q \in \mathbb{N}$, $0 < \epsilon \leq 1$, $\lambda > \mu_q^* > 0$, there is a non-negative power q polynomial $f(x)$ and independent 0/1 random variables X_1, \dots, X_m such that*

- $\mu_j(f, X) \leq \epsilon^{q-j} \mu_q^*$ for all $0 \leq j \leq q$.
- $\Pr [f(X) - \mathbb{E}[f(X)] \geq \lambda] \geq \exp \{-2\epsilon\} \left(\frac{\epsilon}{4q(\lambda/\mu_q^*)^{1/q}} \right)^{4q(\lambda/\mu_q^*)^{1/q}}$.
- $\Pr [f(X) - \mathbb{E}[f(X)] \geq \mu_q^*] \geq \exp \{-2\epsilon\} \left(\frac{\epsilon}{q+1} \right)^{q+1}$.

Proof. We pick $f(x) = \mu_q^* \cdot \sum_{I \subseteq M, |I|=q} \prod_{i \in I} x_i$ where $|M| = m = \lceil 4q(\lambda/\mu_q^*)^{1/q} \rceil$ and each X_i is 1 with probability $\epsilon/m \leq 1/2$. By Lemma 8.2 we have

$$\mu_j(f) \leq \left(m \cdot \frac{\epsilon}{m} \right)^{q-j} \mu_q^* = \epsilon^{q-j} \mu_q^*.$$

The third part of the lemma follows from Lemma 8.2 with $c = q+1$. Indeed $\binom{q+1}{q} = q+1 \geq 2 \geq \epsilon^q + 1 \geq (\mathbb{E}[f] + \mu_q^*)/\mu_q^*$. Therefore,

$$\Pr [f(X) - \mathbb{E}[f(X)] \geq \mu_q^*] \geq \Pr \left[f(X) = \mu_q^* \binom{c}{q} \right] \geq e^{-2\epsilon} \left(\frac{\epsilon}{c} \right)^c = \exp \{-2\epsilon\} \left(\frac{\epsilon}{q+1} \right)^{q+1}.$$

Towards proving the second part of the lemma choose c such that

$$\binom{c-1}{q} \leq \frac{\lambda + \mathbb{E}[f]}{\mu_q^*} < \binom{c}{q}.$$

We have $\lambda + \mathbb{E}[f] \geq \mu_q^*$ so such a $c \geq q+1$ exists. Note that

$$\frac{(c-1)^q}{q^q} \leq \binom{c-1}{q} \leq \frac{\lambda + \mathbb{E}[f]}{\mu_q^*} \leq \frac{\lambda + \mu_q^*}{\mu_q^*} \leq \frac{2\lambda}{\mu_q^*}$$

hence

$$c \leq 1 + q(2\lambda/\mu_q^*)^{1/q} \leq 4q(\lambda/\mu_q^*)^{1/q} \leq m.$$

By the second part of Lemma 8.2 we have

$$\Pr[f(X) - \mathbb{E}[f(X)] \geq \lambda] \geq \Pr\left[f(X) = \mu_q^* \binom{c}{q}\right] \geq e^{-2\epsilon} \left(\frac{\epsilon}{c}\right)^c \geq e^{-2\epsilon} \left(\frac{\epsilon}{4q(\lambda/\mu_q^*)^{1/q}}\right)^{4q(\lambda/\mu_q^*)^{1/q}}.$$

■

The *binomial distribution* $B(n, p)$ is the distribution of the sum of n independent 0/1 random variables each with mean p . The following lower bound on concentration of binomially distributed random variables is well known (e.g. [23] has more general and precise bounds) but we include a proof for completeness in the Appendix.

Lemma 8.4 *For any $\mu \geq 27$ and $0 < \lambda \leq \mu$ there exists a binomially distributed random variable Z with $\mathbb{E}[Z] = \mu$ and*

$$\Pr[Z \geq \mathbb{E}[Z] + \lambda] \geq e^{-100 - \frac{\lambda^2}{\mathbb{E}[Z]}}. \quad (8.42)$$

Remark: The restriction that μ is bounded away from zero is needed since when $\mu = \lambda \ll 1$ the right hand side of (8.42) is constant and the left hand side is necessarily small because $\Pr[Z \geq \mathbb{E}[Z] + \lambda] = \Pr[Z \geq 1] \leq \mathbb{E}[Z] = \mu$.

Lemma 8.5 *For any $q \in \mathbb{N}$, $\mu_q^* > 0$, $\mu_0^* \geq 27\mu_q^*$, $0 < \lambda \leq \mu_0^*$ and $0 < \epsilon \leq 1$ there is a polynomial f of power q and independent 0/1 random variables X_1, \dots, X_m such that*

- $\mu_0(f) = \mu_0^*$,
- $\mu_q(f) = \mu_q^*$,
- $\mu_j(f) \leq \epsilon\mu_q^*$ for all $1 \leq j \leq q-1$,
- $\Pr[f(X) - \mathbb{E}[f(X)] \geq \lambda] \geq e^{-100} e^{-\frac{\lambda^2}{\mu_0^* \mu_q^*}}.$

Proof. We fix a sufficiently large integer n such that $\left(\frac{\mu_0^*}{n\mu_q^*}\right)^{1/q} \leq \epsilon$ and pick our polynomial $f(X)$ to be essentially a linear function in disguise:

$$f(X) = \mu_q^* \cdot \sum_{0 \leq i \leq n-1} X_{qi+1} \cdot X_{qi+2} \cdot \dots \cdot X_{qi+q}$$

where X_i are boolean random variables with $\Pr[X_i = 1] = \left(\frac{\mu_0^*}{n\mu_q^*}\right)^{1/q} \leq \epsilon$. Observe that $\mu_q(f) = \mu_q^*$, $\mu_0(f) = \mu_0^*$, and for $1 \leq i \leq q-1$ we have $\mu_i(f) = \left(\frac{\mu_0^*}{n\mu_q^*}\right)^{(q-i)/q} \mu_q^* \leq \epsilon^{q-i} \mu_q^* \leq \epsilon \mu_q^*$.

Observe that $f(X)/\mu_q^*$ has the same distribution as a binomially distributed random variable with mean $\mu_0^*/\mu_q^* \geq 27$. The lower-bound on $\Pr[f(X) - \mathbb{E}[f(X)] \geq \lambda]$ therefore follows from Lemma 8.4 for sufficiently large n . ■

The following lemma shows how to use a counterexample polynomial of power less than q in place of a counterexample of power q .

Lemma 8.6 For any $\epsilon > 0$, q -uniform hypergraph $H = (\mathcal{V}, \mathcal{H})$, non-negative weights w , polynomial $f(x) = \sum_{h \in \mathcal{H}} w_h \prod_{v \in h} x_v$, independent 0/1 random variables X_1, \dots, X_n and any $q' > q$ there exists a q' -uniform hypergraph $H = (\mathcal{V}', \mathcal{H}')$, non-negative weights w' , independent 0/1 random variables $X'_1, \dots, X'_{n'}$ and polynomial $f'(x') = \sum_{h' \in \mathcal{H}'} w'_{h'} \prod_{i \in h'} x'_i$ such that

- $\mu_i(f') \leq \mu_i(f)$ for all $i \leq q$,
- $\mu_i(f') \leq \epsilon \mu_q(f)$ for all $q < i \leq q'$,
- $\Pr[f'(X') - \mathbb{E}[f'(X')] \geq \lambda] \geq 2^{-(q'-q)} \Pr[f(X) - \mathbb{E}[f(X)] \geq \lambda]$.

Proof. We let $f'(X, Y) = f(X)g(Y)$ where $g(Y)$ is a power- $(q' - q)$ polynomial that is well concentrated around 1. In particular we use

$$g(Y) = \left(\frac{2}{m} \sum_{i=1}^m Y_{1,i} \right) \cdot \left(\frac{2}{m} \sum_{i=1}^m Y_{2,i} \right) \cdot \dots \cdot \left(\frac{2}{m} \sum_{i=1}^m Y_{q'-q,i} \right)$$

where the $n' = (q' - q)m$ random variables Y_{ij} are independent with mean 1/2 and

$$m = \left\lceil \max \left(2/\epsilon, 2 \max_{i,j} \frac{\mu_j(f, X)}{\mu_i(f, X)} \right) \right\rceil.$$

Note that $2/m \leq \epsilon$ and $2/m \leq \frac{\mu_j(f, X)}{\mu_i(f, X)}$ for any $0 \leq i, j \leq q$. It is easy to see that $\mu_i(g, Y) = (2/m)^i$. Let $X' = X_1, \dots, X_n, Y_{1,1}, \dots, Y_{q'-q,m}$ denote the random variables that f' is a function of. By Lemma 8.1 we get that

$$\begin{aligned} \mu_i(f', X') &= \max_{0 \leq j \leq q: 0 \leq i-j \leq q'} \mu_j(f) \mu_{i-j}(g) \\ &= \max_{0 \leq j \leq q: 0 \leq i-j \leq q'} \mu_j(f, X) (m/2)^{-(i-j)} \\ &= \max_{0 \leq j \leq \min(q, i)} B_{ij} \end{aligned} \tag{8.43}$$

where $B_{ij} = \mu_j(f, X) (m/2)^{-(i-j)}$. We bound (8.43) in two cases. The first case is $i \leq q$. For any $0 \leq j < i \leq q$ we have

$$B_{ij} = \mu_j(f, X) (m/2)^{-(i-j)} \leq \mu_j(f, X) (2/m) \leq \mu_i(f, X). \tag{8.44}$$

Clearly $B_{ij} \leq \mu_i(f, X)$ holds for $i = j$ as well, so we conclude that $\max_{0 \leq j \leq \min(q, i)} B_{ij} \leq \mu_i(f, X)$ when $i \leq q$.

The other case is when $i > q$. For any $0 \leq j < q < i$ we have

$$\mu_j(f, X) (m/2)^{-(i-j)} \leq \mu_j(f, X) (2/m)^2 \leq \mu_j(f, X) \cdot \frac{\mu_q(f, X)}{\mu_j(f, X)} \cdot \epsilon = \epsilon \mu_q(f, X). \tag{8.45}$$

Similarly for $0 \leq j = q < i$ we have

$$\mu_j(f, X) (m/2)^{-(i-j)} = \mu_q(f, X) (m/2)^{-(i-q)} \leq \mu_q(f, X) (2/m) \leq \epsilon \mu_q(f, X). \tag{8.46}$$

Combining (8.43), (8.44), (8.45) and (8.46) we conclude that

$$\mu_i(f', X') \leq \begin{cases} \mu_i(f, X), & \text{if } i \leq q, \\ \epsilon \mu_q(f, X), & \text{otherwise.} \end{cases}$$

To show the last part of the lemma we bound

$$\begin{aligned}
\Pr[f(X)g(Y) - \mathbb{E}[f(X)g(Y)] \geq \lambda] &\geq \Pr[f(X) - \mathbb{E}[f(X)] \geq \lambda \text{ and } g(X) \geq 1] \\
&= \Pr[f(X) - \mathbb{E}[f(X)] \geq \lambda] \Pr[g(X) \geq 1] \\
&\geq \Pr[f(X) - \mathbb{E}[f(X)] \geq \lambda] 2^{-(q'-q)}
\end{aligned}$$

where the last inequality follows because each of the linear terms $(\frac{2}{m} \sum_{i=1}^m Y_{j,i})$ in the definition of $g(Y)$ is distributed symmetrically about its mean of 1 and hence is at least one with probability at least $1/2$. \blacksquare

Proof. of Theorem 1.3. Fix q , λ and $\{\mu_i^*\}_{0 \leq i \leq q}$. Let i be the dominant term in (1.2), i.e. i minimizes $\min_i(\lambda^2/(\mu_0^* \mu_i^*), (\lambda/\mu_i^*)^{1/i})$. We consider three cases.

The first case is when $\lambda \leq \mu_i^*$. In this case we apply Lemma 8.3 (third part) to get a power i polynomial and then Lemma 8.6 to convert it into a power q polynomial, using $\epsilon = \min_{0 \leq j, j' \leq q} \mu_j^*/\mu_{j'}^*$ for both Lemmas. This yields 0/1 random variables X_1, \dots, X_n and the desired power q polynomial $f(X)$ with $\mu_j(f, X) \leq \mu_j^*$ for $0 \leq j \leq q$ and

$$\begin{aligned}
\Pr[f(X) - \mathbb{E}[f(X)] \geq \lambda] &\geq \Pr[f(X) - \mathbb{E}[f(X)] \geq \mu_i] \\
&\geq 2^{-(q-i)} \exp \left\{ -2\epsilon + (i+1) \ln \left(\frac{\epsilon}{i+1} \right) \right\} \\
&\geq 2^{-q} \exp \left\{ -2 + (q+1) \ln \left(\frac{\epsilon}{q+1} \right) \right\} = \frac{1}{C_1} \\
&\geq \max \left\{ e^{-\left(\frac{\lambda^2}{\mu_0^* \mu_i^*} + 1\right) \log C_1}, e^{-\left(\left(\frac{\lambda}{\mu_i^*}\right)^{1/i} + 1\right) \log C_1} \right\} \quad (8.47)
\end{aligned}$$

where $C_1 = 2^q e^2 ((q+1)/\epsilon)^{q+1}$.

The second case is when $\lambda > \mu_i^*$ and $27\lambda^2/(\mu_0^* \mu_i^*) \geq (\lambda/\mu_i^*)^{1/i}$. We apply Lemma 8.3 (second part) and Lemma 8.6 using $\epsilon = \min_{0 \leq j, j' \leq q} \mu_j^*/\mu_{j'}^*$ for both, yielding independent 0/1 random variables X_1, \dots, X_n and a degree q polynomial $f(X)$ with $\mu_j(f, X) \leq \mu_j^*$ for $0 \leq j \leq q$ and

$$\begin{aligned}
\Pr[f(X) - \mathbb{E}[f(X)] \geq \lambda] &\geq 2^{-(q-i)} \exp \left\{ -2\epsilon + 4i(\lambda/\mu_i^*)^{1/i} \ln \left(\frac{\epsilon}{4i(\lambda/\mu_i^*)^{1/i}} \right) \right\} \\
&\geq \frac{1}{C_2} (C_3)^{-(\lambda/\mu_i^*)^{1/i}} \\
&\geq \frac{1}{C_2} (C_4)^{-\min((\lambda/\mu_i^*)^{1/i}, \lambda^2/(\mu_0^* \mu_i^*))} \\
&\geq \max \left\{ e^{-\left(\frac{\lambda^2}{\mu_0^* \mu_i^*} + 1\right) \log(\max\{C_2, C_4\})}, e^{-\left(\left(\frac{\lambda}{\mu_i^*}\right)^{1/i} + 1\right) \log(\max\{C_2, C_4\})} \right\} \quad (8.48)
\end{aligned}$$

for $C_2 = e^{2q}$, $C_3 = (\lambda/\mu_i^*)^4 (\frac{4i}{\epsilon})^{4i}$, and $C_4 = C_3^{27}$.

The final case is when $\lambda > \mu_i^*$ and $27\lambda^2/(\mu_0^* \mu_i^*) < (\lambda/\mu_i^*)^{1/i}$. These constraints imply that $\mu_0^* > 27\lambda^{2-1/i} (\mu_i^*)^{1/i-1} \geq 27\lambda$, hence $\lambda < \mu_0^*$. We also have $\mu_0^* > 27\lambda > 27\mu_i^*$. We apply Lemmas 8.5

and Lemma 8.6 with $\epsilon = \min_{0 \leq j, j' \leq q} \mu_j^* / \mu_{j'}^*$ for both, yielding a polynomial f and independent 0/1 random variables X_1, \dots, X_n with $\mu_j(f, X) \leq \mu_j^*$ for $0 \leq j \leq q$ and

$$\begin{aligned} \Pr[f(X_1, \dots, X_n) \geq \mathbb{E}[f] + \lambda] &\geq e^{-100} 2^{-(q-i)} e^{-\frac{\lambda^2}{\mu_0 \mu_i}} \\ &\geq \frac{1}{C_5} e^{-\frac{\lambda^2}{\mu_0 \mu_i}} \\ &\geq \max \left\{ e^{-\left(\frac{\lambda^2}{\mu_0^* \mu_i^*} + 1\right) \log C_5}, e^{-\left(\left(\frac{\lambda}{\mu_i^*}\right)^{1/i} + 1\right) \log C_5} \right\} \end{aligned} \quad (8.49)$$

where $C_5 = e^{100} \cdot 2^q$. This completes the case analysis.

Let $C = \max\{C_1, C_2, C_4, C_5\} \leq c_0 \Lambda_1^{c_1} \Lambda_2^{c_2} \Lambda_3^{c_3}$ for appropriate absolute constants c_0, c_1, c_2 and c_3 where $\Lambda_1 = \max_{0 \leq i, j \leq q} (\mu_i^* / \mu_j^*)^q = \epsilon^{-q}$, $\Lambda_2 = \max_{1 \leq i \leq q} \lambda / \mu_i^*$ and $\Lambda_3 = q^q$. The Theorem follows from (8.47), (8.48) and (8.49). \blacksquare

Acknowledgments

We would like to thank an anonymous referee for many insightful and helpful comments.

References

- [1] S. Aaronson and A. Arkhipov, The Computational Complexity of Linear Optics, in Proceedings of STOC 2011.
- [2] R. Adamczak, Logarithmic Sobolev Inequalities and Concentration of Measure for Convex Functions and Polynomial Chaoses, Bulletin Polish Acad. Sci. Math. 53 (2005), 221-238.
- [3] R. Adamczak and R. Latała, Tail and moment estimates for chaoses generated by symmetric random variables with logarithmically concave tails, to appear in Annales de l'Institut Henri Poincaré, arXiv:1007.1431v1.
- [4] N. Alon and J. Spencer, The Probabilistic Method, third edition, Wiley, 2008.
- [5] M. An, Log-concave Probability Distributions: Theory and Statistical Testing, in Game Theory and Information EconWPA 9611002 (1996).
- [6] M. An, Log-concave Probability Distributions: Theory and Statistical Testing, J. Econometric Theory 80, 350-369.
- [7] M. Arcones and E. Giné, On decoupling, series expansions and tail behavior of chaos processes, J. Theoretical Probability 6 (1993), pp. 101-122.
- [8] M. Bagnoli and T. Bergstrom, Log-concave probability and its applications, Economic Theory 26(2) (2005), pp. 445-469.
- [9] A. Barvinok, Matrices with prescribed row and column sums, to appear in Linear Algebra and its Applications.
- [10] I. Ben-Eliezer, S. Lovett, A. Yadin, Polynomial Threshold Functions: Structure, Approximation and Pseudorandomness, arXiv:0911.3473v3.

- [11] S. Bernstein, On a modification of Chebyshev's inequality and of the error formula of Laplace, vol. 4, N. 5 (original publication: Ann. Sci. Inst. Sav. Ukraine, Sect. Math. 1, 1924).
- [12] C. Borell, On the Taylor series of a Wiener polynomial, In Seminar Notes on Multiple Stochastic Integration, Polynomial Chaos and Their Integration, Case Western Reserve Univ. Cleveland (1984).
- [13] S. Boucheron, O. Bousquet, and G. Lugosi, Concentration inequalities, in O. Bousquet, U.v. Luxburg, and G. Rätsch (editors), Advanced Lectures in Machine Learning, Springer, pp. 208–240, 2004.
- [14] S. Boucheron, O. Bousquet, G. Lugosi and P. Massart, Moment inequalities for functions of independent random variables, *Annals of Probability*, 33(2) (2005), pp. 514-560.
- [15] S. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.
- [16] A. Carbery and J. Wright, Distributional and L^q norm inequalities for polynomials over convex bodies in \mathbb{R}^n . *Math. Res. Lett.* 8 (2001), no. 3, 233-248.
- [17] F. Chung and L. Lu, Concentration inequalities and martingale inequalities: a survey, *Internet Math.* 3 (2006), no. 1, 79-127.
- [18] I. Diakonikolas, P. Harsha, A. Klivans, R. Meka, P. Raghavendra, R. Servedio and L. Tan, Bounding the average sensitivity and noise sensitivity of polynomial threshold functions, In Proceedings of the 42nd ACM symposium on Theory of computing, STOC '10, pp. 533-542, New York, NY, USA, 2010. ACM.
- [19] I. Diakonikolas, P. Raghavendra, R. Servedio and L. Tan, Average sensitivity and noise sensitivity of polynomial threshold functions, *CoRR*, abs/0909.5011, 2009.
- [20] I. Diakonikolas, R. Servedio, L. Tan and A. Wan, A regularity lemma, and low-weight approximators, for low-degree polynomial threshold functions, In Proceedings of the 2010 IEEE 25th Annual Conference on Computational Complexity, CCC '10, pp. 211-222, Washington, DC, USA, 2010. IEEE Computer Society.
- [21] I. Diakonikolas, P. Gopalan, R. Jaiswal, R. Servedio and E. Viola, Bounded independence fools halfspaces, *SIAM J. Comput.* 39(8) (2010), pp. 3441-3462, preliminary version appeared in Proceedings of the FOCS2009, pp. 171-180.
- [22] D. Dubhashi and A. Panconesi, *Concentration of measure for the analysis of randomized algorithms*, Cambridge University Press, Cambridge, 2009.
- [23] W. Feller, Generalization of a probability limit theorem of Cramér, *Trans. Amer. Math. Soc.* 54, (1943), pp. 361-372.
- [24] S. Janson, *Gaussian Hilbert spaces*. Cambridge Tracts in Mathematics. 129. Cambridge: Cambridge University Press, (1997).
- [25] S. Janson, Poisson approximation for large deviations, *Random Structures and Algorithms*, 1 (1990), pp. 221-230.
- [26] S. Janson and A. Rucinski, The infamous upper tail, *Random Structures and Algorithms*, 20(3) (2002), pp. 317-342.

- [27] N. Johnson and S. Kotz, Discrete Distributions, John Wiley, New York (1969).
- [28] D. Hanson and F. Wright, A bound on tail probabilities for quadratic forms in independent random variables, *Ann. Math. Statistics* 42 (1971), pp. 1079-1083.
- [29] P. Harsha, A. Klivans and R. Meka, Bounding the sensitivity of polynomial threshold functions, *CoRR*, abs/0909.5175, 2009.
- [30] D. Kane, The gaussian surface area and noise sensitivity of degree-d polynomial threshold functions, In *Proceedings of the 2010 IEEE 25th Annual Conference on Computational Complexity, CCC '10*, pp. 205-210, Washington, DC, USA, 2010. IEEE Computer Society.
- [31] J. Kim and V. Vu, Concentration of multivariate polynomials and its applications, *Combinatorica* 20 (2000), no. 3, 417-434.
- [32] R. Latala, Estimates of moments and tails of Gaussian chaoses, *Ann. Probab.* 34 (2006), pp. 2315-2331.
- [33] R. Latala and R. Łochowski, Moment and tail estimates for multidimensional chaoses generated by positive random variables with logarithmically concave tails, *Progr. Probab.* 56 (2003), 77-92.
- [34] M. Ledoux, The concentration of measure phenomenon. *Mathematical Surveys and Monographs* 89, American Mathematical Society (2001). Second printing (2005).
- [35] J. Lehec, Moments of the Gaussian Chaos, *Séminaire de Probabilités XLIII, Lecture Notes in Mathematics* 2011, Volume 2006/2011, pp. 327-340.
- [36] P. Major, Estimates on the tail behavior of Gaussian polynomials. The discussion of a result of Latala, *arXiv:0912.2279v1*.
- [37] C. Mathieu and W. Schudy, Correlation clustering with noisy input, In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms* (2010) pp. 712-728.
- [38] R. Meka and D. Zuckerman. Pseudorandom generators for polynomial threshold functions. In *Proceedings of the 42nd ACM symposium on Theory of computing, STOC '10*, pp. 427-436, New York, NY, USA, 2010. ACM.
- [39] C. McDiarmid, Concentration, In *Probabilistic Methods for Algorithmic Discrete Mathematics*, M. Habib, C. McDiarmid, J. Ramirez-Alfonsin and B. Reed editors, pp. 195-248, Springer, 1998.
- [40] M. Mitzenmacher and E. Upfal. *Probability and Computing*, Cambridge University Press, 2005.
- [41] F. Nazarov, M. Sodin and A. Volberg, The geometric Kannan-Lovsz-Simonovits lemma, dimension-free estimates for the distribution of the values of polynomials, and the distribution of the zeros of random analytic functions. (Russian) *Algebra i Analiz* 14 (2002), no. 2, 214-234; translation in *St. Petersburg Math. J.* 14 (2003), no. 2, 351-366.
- [42] R. O'Donnell, *Lecture Notes on Boolean Analysis*, manuscript available online.
- [43] R. Prabhakar Raghavan and C. Thompson, Randomized rounding: a technique for provably good algorithms and algorithmic proofs. *Combinatorica* 7(4): 365-374 (1987).

- [44] M. Talagrand, A new look at independence, *Annals of Probability* 24 (1996), pp. 1-34.
- [45] M. Talagrand, Concentration of Measures and Isoperimetric Inequalities in product spaces, *Publications Mathematiques de l'I.H.E.S.*, 81 (1996), pp. 73-205.
- [46] T. Tao and V. Vu, On the permanent of random Bernoulli matrices, *Advances in Mathematics* 220:3 (2009), pp. 657-669.
- [47] C. Tsourakakis, U. Kang, G. Miller, and C. Faloutsos. DOULION: counting triangles in massive graphs with a coin, in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (2009) pp. 837–846.
- [48] C. Tsourakakis, M. Kolountzakis, and G. Miller. Approximate Triangle Counting, arXiv preprint (2009).
- [49] V. Vu, Concentration of non-Lipschitz functions and applications, *Probabilistic methods in combinatorial optimization. Random Structures Algorithms* 20 (2002), no. 3, 262-316.
- [50] V. Vu, On the concentration of multivariate polynomials with small expectation, *Random Structures Algorithms* 16 (2000), no. 4, 344-363.

A Linear special case

In this section we give a short proof of the linear case ($q = 1$) of Theorem 1.2. Concentration in this case was already known, but this special case nicely illustrates many of our techniques with minimal technical complications. In this case hyperedges are just single vertices, so to simplify notation we make no reference to hyperedges. The rest of the proofs appear in the full version of the paper.

We have n vertices $1, 2, \dots, n$, independent random variables Y_1, \dots, Y_n that are moment bounded with parameter L , and weights w_1, \dots, w_n . We assume that $\mathbb{E}[Y_v] = 0$ and $w_v \geq 0$ for all $v \in [n]$. We are looking for concentration of $f(Y) = \sum_{v \in [n]} w_v Y_v$. Our bounds are based on the parameters $\mu_0 = \sum_{v \in [n]} w_v \mathbb{E}[|Y_v|]$ and $\mu_1 = \max_{v \in [n]} w_v$.

Fix even integer $k \geq 2$. By linearity of expectation and independence we have

$$\begin{aligned} \mathbb{E}[f(Y)^k] &= \sum_{v_1, \dots, v_k \in [n]} w_{v_1} \cdots w_{v_k} \mathbb{E}[Y_{v_1} \cdots Y_{v_k}] \\ &= \sum_{v_1, \dots, v_k \in [n]} w_{v_1} \cdots w_{v_k} \prod_{v \in \{v_1, \dots, v_k\}} \mathbb{E}[Y_v^{|\{i \in [k]: v_i = v\}|}]. \end{aligned} \quad (\text{A.50})$$

For conciseness we write the sum over $v_1, \dots, v_k \in [n]$ in (A.50) as a sum over vectors $\bar{v} \in [n]^k$ (with components v_1, \dots, v_k). The sum over \bar{v} in (A.50) is awkward to bound because it is very inhomogeneous, including e.g. both the case when $v_1 = \dots = v_k$ and the case that all the v_i are distinct. We deal with this issue as follows. Intuitively we generate \bar{v} by first picking the number of distinct vertices $\ell = |\{v_1, \dots, v_k\}| = |\{\bar{v}\}|$, secondly picking a vector $\bar{u} \in [\ell]^k$ (with components $u_1, \dots, u_k \in [\ell]$) of artificial vertices, and finally choosing an injective mapping π from the artificial vertices $[\ell]$ into the real vertices $[n]$ and letting $v_i = \pi(u_i)$. This process generates each vector \bar{v} a total of $\ell!$ times since the names of the artificial vertices are arbitrary. Combining the above with

(A.50) we have

$$\mathbb{E} \left[f(Y)^k \right] = \sum_{\ell=1}^k \frac{1}{\ell!} \sum_{\bar{u} \in [\ell]^k : |\{\bar{u}\}| = \ell} \sum_{\pi \in M(\ell)} w_{\pi(u_1)} \cdots w_{\pi(u_k)} \prod_{u \in [\ell]} \mathbb{E} \left[Y_{\pi(u)}^{|\{i \in [k] : u_i = u\}|} \right] \quad (\text{A.51})$$

where $M(\ell)$ denotes the set of all injective functions from $[\ell]$ to $[n]$. We introduce the notation $d_u = d_u(u_1, \dots, u_k) = |\{i \in [k] : u_i = u\}|$ for the power of $Y_{\pi(u)}$ in (A.51). If any $d_u = 1$ we have $\mathbb{E} \left[Y_{\pi(u)}^{d_u} \right] = 0$, so we can limit the sum in (A.51) to the set $\mathcal{S}_2(\ell)$ of vectors $\bar{u} \in [\ell]^k$ with $d_1(\bar{u}), \dots, d_\ell(\bar{u}) \geq 2$. The constraint that $|\{\bar{u}\}| = \ell$ is clearly satisfied for all $\bar{u} \in \mathcal{S}_2(\ell)$ so we can safely drop it. Note that $\sum_{u \in [\ell]} d_u = k$, so we therefore can reduce the range of ℓ to $1 \leq \ell \leq k/2$. Consequently we have

$$\mathbb{E} \left[f(Y)^k \right] = \sum_{\ell=1}^{k/2} \frac{1}{\ell!} \sum_{\bar{u} \in \mathcal{S}_2(\ell)} \underbrace{\sum_{\pi \in M(\ell)} w_{\pi(u_1)} \cdots w_{\pi(u_k)} \prod_{u \in [\ell]} \mathbb{E} \left[Y_{\pi(u)}^{d_u(\bar{u})} \right]}_{(*)} \quad (\text{A.52})$$

We bound $(*)$ as

$$\begin{aligned} (*) &= \sum_{\pi \in M(\ell)} w_{\pi(u_1)} \cdots w_{\pi(u_k)} \prod_{u \in [\ell]} \mathbb{E} \left[Y_{\pi(u)}^{d_u(\bar{u})} \right] \\ &\leq \sum_{\pi \in M(\ell)} w_{\pi(u_1)} \cdots w_{\pi(u_k)} \prod_{u \in [\ell]} \mathbb{E} \left[|Y_{\pi(u)}|^{d_u(\bar{u})} \right] \\ &\leq \sum_{\pi \in M(\ell)} w_{\pi(u_1)} \cdots w_{\pi(u_k)} \prod_{u \in [\ell]} L^{d_u-1} \cdot d_u! \cdot \mathbb{E} \left[|Y_{\pi(u)}| \right] \\ &\leq \mu_1^{k-\ell} \sum_{\pi \in M(\ell)} \prod_{u \in [\ell]} w_{\pi(u)} L^{d_u-1} \cdot d_u! \cdot \mathbb{E} \left[|Y_{\pi(u)}| \right] \\ &= \mu_1^{k-\ell} L^{k-\ell} \left(\prod_{u \in [\ell]} d_u! \right) \sum_{\pi \in M(\ell)} \prod_{u \in [\ell]} w_{\pi(u)} \mathbb{E} \left[|Y_{\pi(u)}| \right] \end{aligned} \quad (\text{A.53})$$

where the second inequality uses moment boundedness $d_u - 1$ times (per u) and the third inequality follows because $\mu_1 = \max_v w_v$. We now extend the sum over $\pi \in M(\ell)$ in (A.53) (adding additional non-negative terms) to include all mappings from $[\ell]$ into $[n]$ injective or not, which enables us to move the sum over π inside the product over u as follows:

$$\begin{aligned} (*) &\leq (\mu_1 L)^{k-\ell} \left(\prod_{u \in [\ell]} d_u! \right) \sum_{\pi(1), \dots, \pi(\ell) \in [n]} \prod_{u \in [\ell]} w_{\pi(u)} \mathbb{E} \left[|Y_{\pi(u)}| \right] \\ &= (\mu_1 L)^{k-\ell} \left(\prod_{u \in [\ell]} d_u! \right) \prod_{u \in [\ell]} \sum_{v \in [n]} w_v \mathbb{E} \left[|Y_v| \right] \\ &= (\mu_1 L)^{k-\ell} \left(\prod_{u \in [\ell]} d_u! \right) \left(\sum_{v \in [n]} w_v \mathbb{E} \left[|Y_v| \right] \right)^\ell \\ &= (\mu_1 L)^{k-\ell} \left(\prod_{u \in [\ell]} d_u! \right) \mu_0^\ell. \end{aligned} \quad (\text{A.54})$$

Combining (A.52) with (A.54) we get

$$\begin{aligned}
\mathbb{E} \left[f(Y)^k \right] &\leq \sum_{\ell=1}^{k/2} \frac{1}{\ell!} \sum_{\bar{u} \in \mathcal{S}_2(\ell)} (\mu_1 L)^{k-\ell} \left(\prod_{u \in [\ell]} d_u! \right) \mu_0^\ell \\
&= \sum_{\ell=1}^{k/2} \frac{1}{\ell!} (\mu_1 L)^{k-\ell} \mu_0^\ell \sum_{d_1, \dots, d_\ell \geq 2: d_1 + \dots + d_\ell = k} \underbrace{\sum_{\bar{u} \in [\ell]^k: d_1(\bar{u})=d_1, \dots, d_\ell(\bar{u})=d_\ell}}_{(\dagger)} \left(\prod_{u \in [\ell]} d_u! \right) \quad (\text{A.55})
\end{aligned}$$

where the equality groups the sum over \bar{u} by the value of the d_u . Now we claim that the equality $(\dagger) = k!$ follows easily from either an appeal to multinomial coefficients or the following direct argument. Indeed consider k balls of which d_u are labeled u for all $u \in [\ell]$. Each of the $k!$ permutation of the balls induces a vector \bar{u} of the labels. Every $\bar{u} \in [\ell]^k : d_1(\bar{u}) = d_1, \dots, d_\ell(\bar{u}) = d_\ell$ is produced by exactly $\prod_{u \in [\ell]} d_u!$ permutations of the balls, proving the claim.

Substituting $(\dagger) = k!$ into (A.55), bounding the number of different $d_1, \dots, d_\ell \geq 1$ with $d_1 + \dots + d_\ell = k$ by 2^k , and bounding $k!/\ell! \leq R_0^k k^k / \ell^\ell \leq R_0^k \left(\max_{l \geq 1} \frac{k^\ell}{\ell^\ell} \right) k^{k-\ell} \leq R_1^k k^{k-\ell}$ for some constants $1 < R_0 < R_1$ we get

$$\begin{aligned}
\mathbb{E} \left[f(Y)^k \right] &\leq \sum_{\ell=1}^{k/2} \frac{1}{\ell!} (\mu_1 L)^{k-\ell} \mu_0^\ell k! 2^k \\
&\leq (k/2) (2R_1)^k \max_{\ell \in [k/2]} k^{k-\ell} (\mu_1 L)^{k-\ell} \mu_0^\ell \\
&= (k/2) (2R_1)^k \max_{\ell \in [k/2]} (k\mu_1 L)^{k-2\ell} (\sqrt{\mu_0 k \mu_1 L})^{2\ell} \\
&\leq \left(\max\{4R_1 k \mu_1 L, 4R_1 \sqrt{\mu_0 k \mu_1 L}\} \right)^k. \quad (\text{A.56})
\end{aligned}$$

For any $\lambda > 0$ we choose k so that $B \approx \lambda/e$ and apply Markov's inequality, yielding

$$\Pr[|f(Y)| \geq \lambda] \leq e^{-k} \leq e^2 \max\{e^{-\lambda^2/(RL\mu_0\mu_1)}, e^{-\lambda/(RL\mu_1)}\}$$

after some straightforward calculations (see proof of the main Theorem in the full version of the paper) for some absolute constant R .

We now sketch the differences between the above linear case and the general case that is proven in the main body of this paper. In the general case the sequences of vertices v_1, \dots, v_k and u_1, \dots, u_k become sequences of hyperedges. The sums over $u \in [\ell]$ remain sums over vertices.

The biggest conceptual difference in the $q > 1$ case is that we consider the number of connected components in the sequence of hyperedges that replaces u_1, \dots, u_k . Counting the number of sequences of hyperedges with c connected components is substantially trickier than the above bound on (\dagger) .

Bounding the equivalent of $(*)$ by a product of various μ_i is also substantially more involved.

B Proof of the Lemma 8.4

Let Y be Poisson distributed with $\mathbb{E}[Y] = \mu$, i.e. $\Pr[Y = i] = e^{-\mu} \mu^i / i!$ for non-negative integers i . We will first show that Y satisfies (8.42) but with a better constant (96 instead of 100). We will then use a limiting argument to prove the lemma.

Let $\delta = \lambda/\mu \leq 1$ and $\delta' = \delta + \frac{3\sqrt{3}}{\sqrt{\mu}} \leq \delta + 1 \leq 2$. We will frequently use the facts that $0 \leq \delta \leq 1$ and $\frac{3\sqrt{3}}{\sqrt{\mu}} \leq \delta' \leq 2$ without explicit mention. Let $f(t) = \mathbb{E}[e^{tY}]$ and $g_a(t) = f(t)e^{-at}$. We will use Theorem A.2.1 in [4] which states that

$$\Pr[Y > a - u] \geq e^{-tu} [g_a(t) - e^{-\epsilon u} (g_a(t + \epsilon) + g_a(t - \epsilon))] \quad (\text{B.57})$$

for any $a, u, t, \epsilon \in \mathbb{R}$ with $u, t, \epsilon, t - \epsilon$ all positive. We choose these parameters as follows: let $a = (1 + \delta')\mu$, $u = 3\sqrt{3\mu}$, $t = \ln(1 + \delta')$ and $\epsilon = \frac{1}{\sqrt{\mu(1 + \delta')}} \leq \frac{1}{3\sqrt{3} \cdot 1}$. Note that by concavity $t = \ln(1 + \delta') \geq \frac{\delta' \ln(1 + 2)}{2} \geq \frac{3\sqrt{3}}{\sqrt{\mu}} \cdot \frac{\ln 3}{2} > \frac{1}{\sqrt{\mu}} \geq \epsilon$ hence $t - \epsilon$ is positive as required.

A standard calculation (e.g. Lemma 5.3 in [40]) shows that $f(t') = e^{\mu(e^{t'} - 1)}$ and $g_a(t') = e^{\mu(e^{t'} - 1) - at'}$. Therefore

$$\begin{aligned} \ln[g_a(t)] &= \mu(e^{\ln(1 + \delta')} - 1) - \mu(1 + \delta') \ln(1 + \delta') \\ &= \mu(\delta' - (1 + \delta') \ln(1 + \delta')) \\ &\geq -\mu\delta'^2/2 \end{aligned} \quad (\text{B.58})$$

where the inequality follows from applying Taylor's theorem to the function $h(x) = (x - (1 + x) \ln(1 + x))$. We also have

$$\begin{aligned} \ln(g_a(t \pm \epsilon)) &= \mu((1 + \delta')e^{\pm \epsilon} - 1) - (\mu(1 + \delta'))(\ln(1 + \delta') \pm \epsilon) \\ &\leq \mu((1 + \delta')(1 \pm \epsilon + \epsilon^2) - 1) - (\mu(1 + \delta'))(\ln(1 + \delta') \pm \epsilon) \\ &= \mu\delta' - \mu(1 + \delta') \ln(1 + \delta') + \mu(1 + \delta')\epsilon^2 \\ &= \ln[g_a(t)] + 1, \end{aligned} \quad (\text{B.59})$$

where the inequality follows from Taylor's theorem and the fact that $e^{\pm \epsilon} \leq e^{\frac{1}{3\sqrt{3}}} \leq 2$ and the last equality uses the fact that $\mu(1 + \delta')\epsilon^2 = \mu(1 + \delta')(1/\sqrt{\mu(1 + \delta')})^2 = 1$. (Inequality (B.59) is shorthand for two inequalities, one (resp. the other) with $+$ (resp. $-$) substituted for \pm .)

Putting the pieces together we get

$$\begin{aligned} \Pr[Y > (1 + \delta)\mu] &= \Pr[Y > a - u] \\ &\geq e^{-\ln(1 + \delta')3\sqrt{3\mu}} \left[g_a(\ln(1 + \delta')) - e^{-\epsilon 3\sqrt{3\mu}} (g_a(\ln(1 + \delta') + \epsilon) + g_a(\ln(1 + \delta') - \epsilon)) \right] \\ &\geq e^{-\ln(1 + \delta')3\sqrt{3\mu}} g_a(\ln(1 + \delta')) \left[1 - e^{-\epsilon 3\sqrt{3\mu}} 2e \right] \\ &\geq e^{-\delta' 3\sqrt{3\mu}} g_a(\ln(1 + \delta')) \left[1 - e^{-3} 2e \right] \\ &\geq e^{-\delta' 3\sqrt{3\mu}} e^{-\mu\delta'^2/2} \left[1 - e^{-3} 2e \right] \\ &\geq e^{-\delta' 3\sqrt{3\mu} - \mu\delta'^2/2 - 1} \end{aligned} \quad (\text{B.60})$$

where the first inequality uses (B.57), the second inequality uses (B.59), the third inequality uses $\ln(1 + \delta') \leq \delta'$ and $\epsilon 3\sqrt{3\mu} = \frac{1}{\sqrt{\mu(1 + \delta')}} \cdot 3\sqrt{3\mu} \geq 3\sqrt{3/3} = 3$, and the fourth inequality uses (B.58).

Finally we bound

$$\begin{aligned} \delta' 3\sqrt{3\mu} + \mu\delta'^2/2 + 1 &= (\delta + 3\sqrt{3/\mu})3\sqrt{3\mu} + \mu(\delta + 3\sqrt{3/\mu})^2/2 + 1 \\ &= 3\sqrt{3}x + 27 + x^2/2 + 3\sqrt{3}x + 27/2 + 1 \\ &\leq x^2/2 + 6\sqrt{3}x + 42 \\ &\leq x^2 + (6\sqrt{3})^2/2 + 42 = \delta^2\mu + 96 \end{aligned} \quad (\text{B.61})$$

where $x = \delta\sqrt{\mu}$. Equations (B.60) and (B.61) imply that

$$\begin{aligned}\Pr[Y \geq \mathbb{E}[Y] + \lambda] &\geq \Pr[Y > (1 + \delta)\mu] \\ &\geq e^{-96 - \frac{\lambda^2}{\mu}}.\end{aligned}\tag{B.62}$$

To complete the proof of the lemma we use a limiting argument. Let Z_1, Z_2, \dots be random variables where Z_n has binomial distribution $B(n, \mu/n)$. Straightforward calculation (e.g. Theorem 5.5 in [40]) shows that $\lim_{n \rightarrow \infty} \Pr[Z_n = i] = \Pr[Y = i]$ for any integer i (i.e. Z_n converges in distribution to Y). It follows that $\lim_{n \rightarrow \infty} \Pr[Z_n \geq i] = 1 - \lim_{n \rightarrow \infty} \sum_{j=0}^{i-1} \Pr[Z_n = j] = 1 - \sum_{j=0}^{i-1} \Pr[Y = j] = \Pr[Y \geq i]$. Consequently (choose $i = \lceil \mu + \lambda \rceil$) there exists $n' \geq 0$ such that $|\Pr[Y \geq \mu + \lambda] - \Pr[Z_{n'} \geq \mu + \lambda]| \leq |e^{-96 - \frac{\lambda^2}{\mu}} - e^{-100 - \frac{\lambda^2}{\mu}}|$. Combining this fact with (B.62) yields $\Pr[Z_n \geq \mu + \lambda] \geq e^{-100 - \frac{\lambda^2}{\mu}}$, i.e. $Z = Z_{n'}$ satisfies (8.42).